Basins at risk – Predicting international river basin conflict and cooperation

Manuscript submitted to *Global Environmental Change* (November 19, 2012)

Thomas Bernauer[a] and Tobias Böhmelt[b]

[a] Corresponding author
ETH Zurich
Center for Comparative and International Studies and Institute for Environmental Decisions
Haldeneggsteig 4, 8092 Zurich
Switzerland
Phone: +41 44 632 67 71
Fax: +41 44 632 12 89
E–mail: thbe0520@ethz.ch

[b] ETH Zurich
Center for Comparative and International Studies, and Institute for Environmental Decisions
Haldeneggsteig 4, 8092 Zurich
Switzerland
Phone: +41 44 632 93 42
Fax: +41 44 632 12 89
E–mail: tobias.boehmelt@ir.gess.ethz.ch

ABSTRACT

Rapidly increasing water demand combined with supply constraints is widely believed to raise the risk of violence over scarce water resources. Because many of the world's freshwater systems extend across national boundaries, this implies an increased potential for international water conflicts. However, which factors make it actually more likely that we observe conflict? Also, is it possible that cooperation emerges? And, most importantly, which freshwater systems are most likely to be affected by conflict or cooperation? The authors apply in–sample and out–of–sample predictions to new event data for *predicting* and *forecasting* international river basin conflict and cooperation. This allows to examine the predictive power of statistical models and to identify how conflict–prone particular international river basins are – and are likely to be in the future. The paper's results are set in contrast with the findings of an earlier profiling study on the "basins at risk." While similarities are given, substantial differences do exist for which the authors provide explanations.

Keywords: Conflict; Cooperation; Forecasting; International River Basins; Prediction

Conflict of interest: The authors declare "no conflict of interest."

Acknowledgements: The authors acknowledge funding by the European Community 7th Framework Program under the collaborative project "Climate Change, Hydro-conflicts and Human Security" (CLICO), project number 244443.

1. Introduction

Over the past few decades, a substantive scientific literature has emerged that offers important insights into the factors that might influence international river basin cooperation and conflict (e.g. Dinar and Dinar, 2003; Wolf et al., 1999, 2003a, 2003b; Yoffe et al., 2003, 2004; Espey and Towfique, 2004; Furlong et al., 2006; Conca et al., 2006; Brochmann and Gleditsch, 2006, 2009; Gleditsch et al., 2006; Hensel et al., 2008; Zeitoun and Mirumachi, 2008; Gerlak and Grant, 2009; Stinnett and Tir, 2009; Tir and Ackermann, 2009; Zawahri and Gerlak, 2009; De Stefano et al., 2010; Dinar et al., 2010; Zeitoun et al., 2010; Brochmann, 2012).[1] While this literature has developed influential theoretical frameworks that shed light on the underlying mechanisms of conflict or cooperation, and has empirically tested these frameworks either qualitatively or quantitatively,[2] the empirical evidence accumulated on the factors shaping the potential for conflict risk and, conversely, cooperation is primarily *ex–post*. In other words, despite important insights, there are substantial limitations that are associated with these kinds of models and primarily with quantitative studies: statistically significant coefficients may not necessarily tell us much about the actual influence of a given explanatory factor. As Ward et al. (2010: 364) note, policy prescriptions cannot be "based on statistical summaries of probabilistic models."

Recent studies tried to depart from these *ex–post* approaches and relied on combinations of hydrological and economic modeling in computational simulations to examine particular factors that affect the (future) potential for international water cooperation and conflict (e.g. Alcamo et al., 2007; Beck and Bernauer, 2010; Bernauer and Siegfried, 2012; Vörösmarty et al., 2010). However, even these studies cannot analyze the *predictive power* of existing

[1] For a recent overview, see, e.g. Bernauer and Kalbhenn (2010).

[2] An alternative approach is to conceptualize a study's dependent variable in terms of armed conflict between riparian countries and to identify whether sharing a river basin increases the risk of armed conflict (e.g. Gleditsch et al., 2006; Furlong et al., 2006). However, the existing literature does not offer any concrete examples of militarized interstate disputes triggered by or associated with transboundary water problems.

explanatory models and we still lack quantitative work that might *forecast* international river basin conflict and cooperation.[3] This shortcoming seems somewhat surprising, since predictive power is usually regarded as the "gold standard" for assessing the quality of explanatory models in most disciplines, including the social sciences (Goertz, 2006).[4] In the words of Schneider et al. (2010: 1): "anticipating the future is both a social obligation and intellectual challenge that no scientific discipline can escape."

In the following paper, we seek to address this gap and revisit the "basins at risk" study originally carried out by Yoffe et al. (2003). This work relies on the transboundary freshwater dispute database (Wolf et al., 1999, 2003a, 2003b; in the following referred to as "TFDD"), which is comprised of event data on international river conflict and cooperation along a continuum, i.e. the Basins at Risk (BAR) scale. Yoffe et al. (2003) regressed this scale on a wide range of factors that might affect conflict risk or the chances of cooperation. Based on their binary regression models, these scholars then identified a number of international water catchments that appeared particularly risk–prone from according to various covariates[5] and categorized these basins along three categories: a) basins in which water conflict was already manifest; b) basins in which conflict is possible in the future and for which there is evidence of existing tensions; and c) basins in which conflict is possible in the future, but there is no present evidence of existing tensions. Table 1 aggregates these categories and summarizes those basins that Yoffe et al. (2003: 1123) predicted as "being at risk."

---

[3] A forecast "is a conditional statement about how a phenomenon will develop in the future" (Bechtel and Leuffen, 2010: 309). In other words, we only use the term forecast when we want to predict a variable "whose values are truly unknown" (Bechtel and Leuffen, 2010: 311). Conversely, predictions as such are "conditional statements about a phenomenon for which the researcher actually has data, i.e. the outcome (or dependent) variable has been observed" (Bechtel and Leuffen, 2010: 311). Note that the term "out–of–sample prediction," which we refer to below, might pertain to both forecasting and prediction in the traditional sense.

[4] For an overview about those scholarly articles dealing with prediction and forecasting, particularly in international relations and with regards to conflict and cooperation, Gleditsch and Ward (2013: 1) suggest, e.g. Choucri and Robinson (1978), Singer and Wallace (1979), Beck et al. (2000), Ward et al. (2007), Goldstone et al. (2010), or O'Brian (2010). See also Schneider et al. (2010, 2011) and Bueno de Mesquita (2011).

[5] In their analysis, the following factors turned out to point to international river basins particularly risk–prone: high population density, low per capita GDP, overall unfriendly relations between riparian countries, politically active minority groups that might internationalize water conflicts, proposed large dams or other water development projects, and limited or no freshwater treaties.

Table 1 in here

We consider such a risk profiling of international river basins not only as scientifically valuable, but also as practically highly useful. Identifying those basins that are likely to be particularly prone to conflict or cooperation is of high interest to policy makers, non–governmental organizations, or international organizations. Hence, while this paper builds upon the basic idea of predicting basins at risk, we develop an approach that addresses this issue within a methodologically more rigorous predictive framework. The paper proceeds as follows. The next section briefly outlines the rationale behind our work and, ultimately, the motivation to move beyond *ex–post* statistical inference and towards systematic predictions and forecasting. Afterwards, we present the underlying empirical background for our succeeding in– and out–of–sample predictions and forecasting techniques. The discussion of our results is at the core of the following section, while we finish the paper with a summary of the results in light of the earlier risk–profiling exercise by Yoffe et al. (2003).

2. From statistical inference to prediction and forecasting

Several scholars argue that drawing inferences from statistically significant results can be misleading to the extent that those inferences are unlikely to tell us much about the predictive power of a specific covariate or an entire model (e.g. Ward et al., 2010; Gleditsch and Ward, 2013). Hence, statistically significant results may improve our understanding of the relationship between variables in a given sample under study, but they may not provide information on the exact same relationship in another, i.e. new sample of data – like the future. Yoffe et al. (2003:1123) are at least discreetly aware of this limitation when stating that "[c]ategorizing a basin at risk does not presume to identify basins in which acute conflict

will occur, but to point to basins worth more detailed investigation." Nonetheless, the ability to predict and forecast which international river basins are more likely to experience conflict or cooperation is of great interest to academics and policy makers alike. In the words of Ward et al. (2010: 364):

"basing policy prescriptions on statistical summaries of probabilistic models (which are predictions) can lead to misleading policy prescriptions if out–of–sample predictive heuristics are ignored. Often such policy prescriptions might be based on factors that are important in one sample, but not very salient in the out–of–sample cases to which they are implicitly being applied. It is quite possible to focus on statistically significant results that are artifacts in the sense that they do not generalize beyond the specific cases studied. This happens if we focus only on statistically significant relationships and may actually hinder our ability to generalize to out–of–sample situations, such as the future! [...] We have to be willing to make predictions explicitly – and plausibly be wrong, even appear foolish – because our policy prescriptions need to be undertaken with results that are drawn from robust models that have a better chance of being correct. The whole point of estimating risk models is to be able to apply them to specific cases. [...] Predictive heuristics provide a useful, possibly necessary, strategy that may help scholars and policymakers guard against erroneous recommendations."

Accordingly, we examine a) the predictive power of explanatory models on international freshwater conflict and cooperation that are based on a recent study by Zawahri and Mitchell (2011), and b) the ability of these models to forecast basins at risk and those that are likely to see cooperation in the future. We use real prediction and forecasting methods, i.e. in–sample prediction and out–of–sample forecasting approaches as suggested in Ward et al. (2010; see also Gleditsch and Ward, 2013; Bechtel and Leuffen, 2010). In–sample and out–of–sample techniques allow for an assessment of the predictive capacity of an entire model or of single covariates in explaining the likelihood of conflict and cooperation over international river basins. This seems important not only for identifying international river basins that are prone to conflict or cooperation at present, but also in terms of the future. Hence, we also employ our techniques for predicting, i.e. forecasting which international river basins are likely to see conflict and/or cooperation over the period 2012–2015.

3. Empirical background: International river basin cooperation and conflict

We start with estimating a model taken from Zawahri and Mitchell (2011; in the following referred to as "Z–M") who also focus on international river basin conflict and cooperation. While our estimations rely on their core empirical model as a baseline for identifying the statistical significance of predictors, we substitute the dependent variables in Z–M for alternative and recently compiled event data (Kalbhenn and Bernauer, 2012; Kalbhenn, 2011). In other words, while the explanatory variables are fully taken from Z–M for our models, we replace their dependent variables (i.e. indicators for multilateral or bilateral river treaties between states) by two alternative indicators that are likely to capture international river basin conflict and cooperation more accurately. These latter variables are outlined in detail below. After estimating these statistical models, we then move beyond existing approaches that exclusively rely on the statistical significance of explanatory variables, since we conduct in–sample and out–of–sample predictions and forecasting. Ultimately, this allows us to identify the most important factors for predicting and forecasting international river basin conflict and cooperation – and basins at risk by implication.

We chose the empirical model of Z–M as a baseline for two reasons. First, their core model operates with the unit of analysis that is of interest to this paper, i.e. yearly observations for an international river basin per country pair (dyad) (in the following referred to "dyad–basin–year"). Second, there is also a high level of methodological sophistication in Z–M due to time–series cross–section nature of the data in 1816–2001, which are based on a careful case selection of rivers (see Owen et al., 2004) and global country coverage as specified in the Correlates of War Project (Singer, 1988).

As noted above, Z–M measure cooperation in the form of treaties. According to these scholars, cooperation over international river basins thus emerges in a specific year when a riparian dyad concludes a treaty (see also, e.g. Espey and Towfique, 2004; Stinnett and Tir,

2009; Tir and Ackermann, 2009; Dinar et al., 2010). The main empirical difference between our model and Z–M is that we use alternative data for the dependent variables, i.e. an event data set on international river basin conflict and cooperation (in the following referred to as "IRCC") between 1997 and 2007 that was compiled by Kalbhenn and Bernauer (2012; see also Kalbhenn, 2011). The primary reason for our choice using event data is that the zero value of a dependent (outcome) variable, which focuses on treaty adoption (1=treaty adopted; 0=no treaty adopted), seems to be a problematic and an arguably inaccurate proxy for conflict onset. For instance, concluding a treaty in a given year may set in motion a cooperative process that lasts for several years. However, "no treaty" in those subsequent years can hardly be considered as an expression of conflict. Similarly, in the absence of significant water scarcity or pollution problems, states may regard a treaty as unnecessary, but the absence of such a treaty should not be treated as an indication of conflict in this case. We therefore believe that defining the dependent variable in terms of treaty adoption is certainly useful for explaining cooperation, but is likely to be too error–prone for capturing the other side of the spectrum of water–related interactions between states, i.e. conflict.[6] At this point, it should be noted that international river basin conflict can, according to the IRCC coding rules, in effect include violent conflict, e.g. militarized interstate disputes over water resources. However, all events in these data are of non–violent nature and, thus, all cases either pertain to international political disputes or international cooperative efforts concerning international river basins.

The explanatory variables in Z–M relate to three theoretical concepts: state interests, transaction costs, and the distribution of power. We briefly summarize these variables here and refer the reader for details and a discussion of the underlying theoretical arguments to Z–

[6] Note that other event data on riparian country interactions in international river basins include the TFDD and the Issue Correlates of War Project (ICOW) by Hensel et al. (2008). On one hand, the latter do not cover all international river basins globally, though. Moreover, parts of the ICOW data are not (yet) publicly available, and their focus is mainly on conflict rather than cooperation. On the other hand, the TFDD offers event data for a longer time period than the IRCC data we use. However, the main reason for using the IRCC data instead of the TFDD is that major changes in the availability of (translated) news media texts over time (notably the advent of the digital revolution) make it problematic to use event data coded from these (changing) sources for a very long period of time. Hence, we prefer to use event data that are based on a more uniform pool of news media reporting. See Kalbhenn and Bernauer (2012) for a comparison of the IRCC and the TFDD.

M (11ff).[7] First, empirical proxies for state interests include the share of a country's surface area located in a given international river basin (*Territory in Basin*), the extent to which a country is dependent on external sources of freshwater (*External Water Dependence*), and an item on precipitation levels, which is another proxy for water availability (*Precipitation*). Z–M use a weakest–link specification for these variables, i.e. the country in a dyad–basin–year that displays the lowest value on a given variable determines the overall value of this item. Yoffe et al. (2003:1116) use these three explanatory variables also in their work.

Transaction costs are measured by the lowest Polity IV (Marshall and Jaggers, 2004) democracy–minus–autocracy score in a dyad (*Democracy*). Somewhat surprisingly from our view, this variable is not considered in Yoffe et al. (2003). Another indicator for transaction costs captures the similarity of domestic legal traditions in a dyad. The categories used to that end are civil law, common law, Islamic law, and mixed law. When two states in a dyad share the same legal tradition, this item (*Legal System Similarity*) is coded as 1 (0 otherwise). The variable in Yoffe et al. (2003) that comes closest to the underlying theoretical rationale pertains to the "overall relations in terms of friendship or hostility" between two states in a dyad. In addition, Z–M include the total number of states in a basin (*Number of Riparian States*) and a variable on direct land–based contiguity (*Contiguity*) in their model. These two items are also considered by Yoffe et al. (2003).

The distribution of power between states in a dyad (*Upstream Power* and *Downstream Power*) is measured in terms of the upstream and downstream countries' economic and military capabilities that are operationalized via CINC scores (Singer, Bremer, and Stuckey, 1972). Yoffe et al. (2003) do not explicitly consider the information on upstream or downstream countries, but these scholars include two relative power measures (the ratio of

[7] We changed the original variables names of Z–M slightly to facilitate the interpretation of results. For a direct comparison, however, see Table 2 below. Also note that Z–M (12f) consider variables for trade dependence, states' memberships in IGOs, and INGOs. We omit these items from our study due to the limited availability of data that severely constrains the number of observations in our (and their) estimated models (see also Yoffe et al., 2003: 1116).

GDP per capita between dyad–basins and the ratio of their population densities) that are differently measured than in Z–M, though. Table 2 summarizes these variables and their sources.

Table 2 in here

The empirical analysis by Z–M produces several consistent and robust findings. In short, stronger dependence on freshwater resources makes cooperation in the form of treaty adoption more likely, while higher precipitation levels make it less likely. Moreover, democratic countries have a higher probability of cooperation; the same positive effect is found for the variables on the similarity of countries' legal systems and geographic contiguity. The results for the other explanatory variables are either inconsistent across bilateral and multilateral river basin treaties or statistically insignificant. Z–M (20) conclude that "except for state interest, the other factors that promote cooperation vary by context." Our reading of their results suggests that democracy, similarity of the legal system, and contiguity are conducive to cooperation regardless of the context (defined by Z–M as bilateral vs. multilateral basin treaties).

4. In–sample predictions: Assessing the predictive power with "existing data"

We now examine the in–sample and out–of–sample predictive power a) of full models on river basins conflict and cooperation; b) of single predictors in these models; and c) by identifying those basins that are most likely to be at risk or see cooperation within the next few years, i.e. the forecasting. We first estimate *ex–post* empirical models that are based on Z–M, but differ in the following specifications. First, we use the IRCC data by Kalbhenn and Bernauer (2012), which we aggregate to the dyad–basin–year, and generate two binary

variables for conflict and cooperation that serve as our dependent variables.[8] The first binary variable (*Conflict*) receives a value of 1 if the median IRCC score for a dyad–basin–year is negative (0 otherwise); the second variable (*Cooperation*) receives a value of 1 if the median IRCC score is positive (0 otherwise). Therefore, we also estimate two separate models: one for conflict (Model 1) and one for cooperation (Model 2). Due to the temporal scope of the IRCC data, our models cover the time period from 1997 to 2001, i.e. that period of time for which both the IRCC dataset and Z–M provide data. While this temporal limitation might appear as a shortcoming of our research at first sight, note that it also offers noticeable opportunities for our out–of–sample predictions, i.e. the forecasting: as we have empirical data for the dependent variables (but not the explanatory variables) in 2002–2007, we are able to use the *ex–post* models for 1997–2001 to predict cooperation and conflict in 2002–2007. Afterwards, we can compare those out–of–sample predictions with the empirically observed values on the dependent variables for the latter period in order to assess the models' forecasting capabilities.

Second, in addition to the new dependent variables, we consider two other components that have not been included in Z–M. On one hand, we incorporate a conflict–years variable and a cooperation–years variable, respectively, as well as different sets of cubic splines to correct for any temporal dependencies (Beck, Katz, and Tucker, 1998). On the other hand, standard errors are clustered on the dyad–basin to control for intra–group correlations. Our results are summarized in Table 3.

Table 3 in here

For Model 2, we find that higher water dependence is associated with more cooperation, which corresponds to the findings in Z–M. Conversely, we might expect a negative

[8] A prerequisite for the prediction and forecasting approaches is a dichotomous dependent variable.

coefficient for the water dependence variable in Model 1. While this coefficient is in fact negative, it is not significant at conventional levels. Model 2 also lines up well with Z–M in that higher precipitation levels make cooperation less likely. However, we also observe a negative impact of precipitation levels on the probability of conflict (Model 1). This finding suggests that there might be a curvilinear relationship between precipitation levels and the continuum of international river basin conflict and cooperation. Democracy has a conflict–reducing effect. In contrast to Z–M, democracy also reduces the probability of cooperation (Model 2), though. Again, we suspect that there is a curvilinear relationship between democracy and the underlying IRCC scale, which covers the entire spectrum of conflict and cooperation. Similarity of the legal system has a significantly negative effect on conflict risk, but in contrast to Z–M, this variable has no significant impact on cooperation. The contiguity variable, which is a strong positive predictor of cooperation in Z–M, drops out in our conflict model and has a negative effect on cooperation in Model 2. Since our dependent variables and time period differ from those in Z–M, these initial results should not be viewed as an empirical contest between two models, data sets, or samples, but rather as a plausibility check of our empirical setup. Either way, prediction and forecasting, which we now move to, are anyway "beyond the issue of the sign and significance of particular coefficients" (Gleditsch and Ward, 2013: 16).

For the in–sample predictions, we group the predicted probabilities of either Model 1 or Model 2 into quintiles and compare these with de facto instances of conflictive or cooperative dyad–basin–years in our data. For these calculations, we refer to the fifth, fourth, and third quintile as the "most–likely" group, i.e. those groups of predicted probabilities that are most likely to match with actual instances of cooperative or conflictive dyad–basin–years. The first and the second quintile, in contrast, are regarded as the "least–likely" group, which is equivalent to those groups of predicted probabilities that are comparatively low and, thus, are

less likely to correctly predict observed onsets of cooperation or conflict.[9] We summarize our findings here in Table 4 and Figure 1.

Table 4 and Figure 1 in here

It turns out that the predictive power of both models is in fact very high. The most–likely group for Model 1 includes 20 out of 21 conflictive dyad–basin–years in our empirical data (95%). With regard to Model 2, 431 out of 468 (92%) cooperative dyad–basin–years are placed in the most–likely group. In other words, only one dyad–basin–year that was de facto conflictive is characterized as a least–likely case. Similarly, only 37 out of 468 cooperative dyad–basin–years are classified as non–cooperative, although they were in fact cooperative. Note that this initial test of the in–sample predictive power appears more promising than, e.g. a similar in–sample prediction exercise recently applied to international armed conflict (Gleditsch and Ward, 2013): when trying to predict militarized interstate disputes, Gleditsch and Ward (2013) find that there were 227 empirically observable dyad–years in which such a dispute occurred. However, their best model estimation is only able to predict 44 out of these 227 cases (19.38%).

Figure 2 sheds more light on the in–sample predictive power of our models in Table 2. The Receiver Operator Characteristic (ROC) plot shows the extent to which models with more predictive power generate "true positives at the expense of fewer false positives" (Ward et al., 2010: 366). Thus, a perfectly predictive model would correctly classify all empirically observed cooperative or conflictive dyad–basin–years and never generate false positives, i.e. dyad–basin–years where predicted cooperation or conflict did not occur although our estimations predict the contrary. Figure 2 emphasizes that our models do not perfectly predict either conflict or cooperation among riparian states in international river basins, yet these

[9] Note that using quintiles instead of, e.g. terciles is arguably more appropriate, since this approach pays more attention to the long tail in the distribution of the predicted probabilities.

models yield a higher predicted probability for a randomly chosen event than for a randomly chosen non–event. This finding is reflected in the area under ROC curve statistic (AUC), which theoretically varies between 0.5 (no predictive power) and 1.0 (perfect predictive power). As indicated in Table 4 and Figure 1 shown above and demonstrated more thoroughly by Figure 2, our models perform well above average in this regard: Model 1 has an AUC value of 0.85 and Model 2 has an AUC statistic of 0.83.

Figure 2 in here

5. Out–of–sample predictions: Forecasting on "new data" and a 4–fold cross–validation quasi–experimental setup

The findings presented in the previous section leave us with the question of how the predictive power of our models looks like when moving to the "harder" test of an out–of–sample prediction, i.e. what is the predictive power of our models when trying to correctly predict outcomes that are not "within the very same set of data that was used to generate the models in the first place" (Ward et al., 2010: 370). Our first step in this context consists again of grouping the predicted probabilities by quintiles and comparing these with the empirically observed conflictive or cooperative dyad–basin–years. The difference between the models behind Table 4/Figure 1 above and those unreported models leading to Table 5/Figure 3 below, however, is that we now use the covariate values in 2001 exclusively for predicting international river basin conflict and cooperation between 2002 and 2007. Z–M and, therefore, also we lack data for our explanatory variables in this time period (not for our dependent variables). Put differently, we use data for the explanatory variables in the last observed year (2001) to predict instances of conflict or cooperation, as measured by our

dichotomous dependent variables, afterwards. Again, we refer to the fifth, fourth, and third quintile as the most–likely group, while the first and the second quintile are designated as the least–likely group.

Table 5 and Figure 3 in here

When employing this harder test, the most–likely group comprises 20 out of 34 conflictive dyad–basin–years in our data (59%). Even more accurately, 821 out of 855 dyad–basin–years (96%) are captured by the cooperative most–likely group as demonstrated in the third column of Table 5 and the right panel in Figure 3. This means that, although we now use "new data," i.e. dyad–basin–years pertaining to the dependent variable that were originally not covered by the models in Table 3, the predictive power (or, more precisely, the quasi–forecasting power in this case of "new" data)[10] has increased in this first step of our out–of–sample predictions relative to the in–sample predictions – at least in the case of cooperative dyad–basin–years. This finding is also mirrored by the values of the ROC plots in Figure 4. The prediction accuracy of the model for conflictive dyad–basin–years increases by 18% to 0.98, while the predictive power of the cooperation model increases by 13% to 0.96.

Differences between the most–likely groups identified in Table 5/Figure 3 and the ROC plots in Figure 4 can be explained by the way we grouped the most–likely observations in the former case. As noted by Gleditsch and Ward (2013: 16), any "threshold for considering an event as predicted could be seen as an arbitrary description of the continuous distribution of the probabilities." Despite our careful selection of the thresholds for the most–likely and the least–likely group along quintiles, this argument by Gleditsch and Ward (2013) clarifies why the ROC curves in Figures 2 and 4 provide a more accurate picture of the predictive power of

[10] We refer to this exercise as quasi–forecasting due to the definitions provided in footnote 3: essentially, we do not analyze "new" data, but treat the existing information as if some parts are "truly unknown" (Bechtel and Leuffen, 2010: 311).

our models. Nevertheless, since the ROC plots in Figure 4 are essentially based on predictions with "new" data, we conclude that the high AUC statistics clearly reveal that our models from Table 3 also seem to have the ability to produce very accurate forecastings.

Figure 4 in here

To further assess the predictive power of our models, we use a 4–fold cross–validation quasi–experimental setup that was repeated 10 times (Ward et al., 2010: 370) – either for the full model (see Table 3) or a model that omits some explanatory variables from the estimation. In more detail, this cross–validation approach relies on dividing the existing data we used for the models in Table 3 into four subsets, while the dyad–basin–year observations are randomly assigned to the different sets. All except one of the subsets are then pooled together and the pooled set of observations are used to estimate the models with their specifications we used in Table 3. The remaining subset, also called the "test set" (Ward et al., 2010: 370), which we did not use for the pooled set of observations, then serves to assess the predictive power of the model estimated on the pooled subsets. Again, we calculate the area under the ROC curve for measuring the predictive power (Ward et al., 2010: 370). Figure 5 (conflictive dyad–basin–years) and Figure 6 (cooperative dyad–basin–years) summarize our findings.

Figures 5 and 6 in here

Despite the fact that the 4–fold cross–validation approach actually relies on observed data, it is even more "conservative" than the out–of–sample strategy used above for the time period 2002–2007. It is therefore not surprising that the predictive power both of the full models and the reduced models, where we exclude some explanatory variables, decreases for the 4–fold

cross–validation. Nevertheless, the predictive power of the full models remains reasonably high: the full model, which includes all explanatory variables from Table 3, scores a 0.75 in predicting conflictive dyad–basin–years and 0.84 in predicting cooperative dyad–basin–years. Note, however, that the average AUC of our estimations where we exclude the control variables consecutively is lower than the AUC of the out–of–sample prediction for any full model. This means that the power of our models to predict and forecast international river basin conflict or cooperation remains unchanged even when implementing the tougher out–of–sample prediction via the 4–fold cross–validation.

Starting with the full models, we dropped individual covariates one–by–one from the two models and examined the implications for the out–of–sample predictive power. The reason for this is simple: demonstrating that an entire model or its alternative specifications perform above the AUC level of 0.5 does not allow for firm conclusions concerning the predictive power of individual covariates therein. We thus follow Ward et al. (2010: 367) and estimate the AUC for each model, dropping covariates one after the other. This approach allows for a direct comparison of the predictive power of each explanatory variable. The stages in Figure 5 and Figure 6, respectively, summarize these results. The interpretation here is also straightforward: those predictors that are associated with lower values either in Figure 5 or 6 have the higher power for predictions and forecasting. This is because the data points in these figures pertain to separate 4–fold cross–validations that *exclude* these predictors. Therefore, for example, *Democracy* is the strongest predictor in the first stage of the conflict model. When dropping this variable in the conflict model, the out–of–sample predictive power decreases from about 0.75 to 0.73. Similarly, in the cooperation model, *Contiguity* (decrease in AUC from 0.84 to about 0.82) is the strongest predictor at stage 1. We then discard both strongest predictors from the respective model for the second stage and repeat the procedure of a 4–fold cross–validations. After identifying the strongest predictors in the second stage, we reiterate this procedure for a third stage. The results for the third stage, for instance, then

indicate that the power of our explanatory variables to predict and forecast international river basin conflict or cooperation is quite high. Our tests show, therefore, that dropping the explanatory variables included in Table 3 from any model estimation would not only be misleading from the perspective of statistical significance, but also from the viewpoint of predictive power. Moreover, Figure 5 and Figure 6 crucially emphasize that more parsimonious models can perform much better for the prediction/forecasting of international river basin conflict or cooperation than more complex models. In essence, already (and only) the three strongest predictors for either conflict or cooperation seem "to do a pretty good job." And given that any of these best predictors is fairly easily measurable or observable, we believe that forecasting of international river conflict or cooperation becomes more accessible in further research.

6. Basins at risk – Revisited

After conducting the in–sample predictions and out–of–sample forecasting estimations, we are now able to compare our findings with the original basins at risk list by Yoffe et al. (2003). Additionally, note that we also estimated unreported models similar to the approach for 2002–2007 above in order to forecast conflictive and cooperative basins for the time period 2012–2015. Ultimately, we aggregated the predicted dyad–basin–years both from the in–sample prediction and all out–of–sample tests to the basin form, i.e. we now discard the dyad–year information from our original unit of analysis and merged the predicted/forecasted observations simply along the basin in order to facilitate comparison with Yoffe et al. (2003). Table 6 summarizes our results.[11]

[11] Table 6 summarizes these basins in a grouped fashion, e.g. we do not highlight those cases that are based on the in–sample prediction. The online appendix, though, presents tables that differentiate between a) basins identified by the in–sample prediction; b) basins identified by the out–of–sample prediction for 2002–2007; c) basins identified by the out–of–sample prediction via the 4–fold cross–validation quasi–experimental setup; and d) basins identified by the out–of–sample prediction for 2012–2015.

Table 6 in here

The first column includes all basins placed by Yoffe et al. (2003) in any of their three risk categories. Hence, this column mirrors Table 1 above. The second column includes all basins that appear in our in–sample and out–of–sample predictions or forecastings of conflict. When comparing the first and second columns in Table 6, we see that our list is slightly shorter, i.e. 25 basins instead of 29. It also turns out that there is a rather limited overlap, since only 12 basins appear in both lists: Aral Sea, Ganges–Brahmaputra–Meghna, Han, Indus, Kunene, Lake Chad, Nile, Ob, Senegal, Tigris–Euphrates, Yalu, Zambezi. With regard to the third column, we obtain strong evidence that it is worth studying international river basin conflict and cooperation side–by–side (see also Zeitoun and Mirumachi, 2008; Zeitoun et al., 2010). This column shows that 16 basins, which are categorized as basins at risk by Yoffe et al. (2003), are also predicted to have at least one dyad–basin–year with a cooperative median.

In other words, while our approach has some basins at risk in common with Yoffe et al. (2003), substantial differences exist. What induces these dissimilarities? Assuming that the IRCC data and the TFDD cover the same concepts (see Kalbhenn and Bernauer, 2012), the dissimilarities have to be caused by the different methodological approaches. However, we believe that our approach leads to the more accurate predictions. As stated above, bivariate regression models in Yoffe et al. (2003) can uncover information about statistical significance. They cannot uncover information about relationships between variables in other samples or with regard to the predictive power, though.

7. Conclusion and discussion

The principle purpose of this research was to put *ex–post* empirical models of international river basin conflict and cooperation to a harder test by examining their power to predict and forecast dyad–basin–years. The empirical models we estimated perform very well in this regard, both for the in–sample tests (1997–2001) and for those out–of–sample approaches that allow for a comparison of the predicted conflictive as well as cooperative dyad–basin–years with empirically observed cases (2002–2007; the 4–fold cross–validation quasi–experimental setup). We thus conclude that the approach taken here is useful for developing more robust explanatory models of transboundary water conflict and cooperation.

Based upon our findings, we identify a potential caveat and suggest further avenues for future research. First, one limitation of our proposed approach might be that it reduces information on conflict and cooperation intensity to binary variables. As indicated above, and elaborated more thoroughly in Kalbhenn (2011) and Kalbhenn and Bernauer (2012), the original IRCC scale ranges from –5 to +6. Our setup of creating dichotomous items in this paper was not only necessary, but also seemed acceptable because very conflictive or extreme cooperative international water events are rare.[12] Similarly, as our empirical setup uses the median value of the IRCC for the dyad–basin–year as the unit of analysis, virtually all extreme conflictive events might have been averaged, in many cases potentially to 0. Still, our dependent variables do aggregate and, thus, drop available information at least to some extent.

Second, further research could examine the predictive power of models that focus on cooperative (rather than conflictive) international river basin events more thoroughly, e.g. the adoption of water–related treaties as in Z–M. Such work would benefit from data availability for longer time periods, although it could arguably not address the basins at risk issue directly.

[12] More specifically, Kalbhenn and Bernauer (2012; see also Kalbhenn, 2011) identify 5,881 events in total. Out of these, only 77 events score an IRCC value of –3 or less (1.31%), while there are only 10 events that score a value of +5 or higher (0.17%).

Recall that this is precisely the reason why we used the IRCC data instead of data on treaty adoptions as in Z–M. Besides, qualitative case studies could study more thoroughly individual river basins to examine some of our findings that may appear counterintuitive at first sight. For instance, the Jordan is listed as a basin at risk in Yoffe et al. (2003), but not identified by any of our in–sample and out–of–sample predictions for 1997–2007. Similarly, it could be insightful to study in detail the false–positive and false–negative river basins, i.e., those cases that were incorrectly classified by our setup. For example, Table 7 lists those false–negative river basins.

Table 7 in here

Finally, further research might want to examine why some basins experience both cooperation and conflict, within specific dyad–basin–years or over time, whereas others are dominated by one of the two interaction types (see also Zeitoun and Mirumachi, 2008; Zeitoun et al., 2010).

Table 1
Basins at risk identified by Yoffe et al. (2003).

Aral Sea	Lempa
Asi/Orontes	Limpopo
Ca	Mekong
Chiloango	Nile
Cross	Ob
Drin	Okavango
Ganges–Brahmaputra–Meghna	Red
Han	Saigon
Indus	Salween
Irrawaddy	Senegal
Jordan	Song Vam Co Dong
Kune	Tigris–Euphrates
Kura–Araks	Yalu
La Plata	Zambezi
Lake Chad	

Table 2
Overview of explanatory variables.

Zawahri and Mitchell (2011)	This paper	Source
% Lowest area in basin	Territory in Basin	International River Basin Registry (Wolf et al., 1999)
Lowest water dependence	External Water Dependence	AQUASTAT Database
Lowest avg. precipitation	Precipitation	AQUASTAT Database
Lowest polity score	Democracy	Polity IV Dataset (Marshall and Jaggers, 2004)
Same legal system	Legal System Similarity	Zawahri and Mitchell (2011)
Contiguity	Contiguity	PRIO Database (UCDP, 2008)
Number of states in basin	Number of Riparian States	Zawahri and Mitchell (2011)
Upstream state CINC	Upstream Power	Correlates of War (Singer, 1988)
Downstream state CINC	Downstream Power	Correlates of War (Singer, 1988)

Table 3
Logistic regression models: International river basin conflict and cooperation

	Model 1 – Conflict	Model 2 – Cooperation
Territory in Basin	0.018	0.015
	(0.012)	(0.005)***
External Water Dependence	–0.010	0.014
	(0.012)	(0.003)***
Precipitation	–0.001	–0.001
	(0.000)**	(0.000)***
Democracy	–0.139	–0.023
	(0.048)***	(0.011)**
Legal System Similarity	–0.911	–0.040
	(0.505)*	(0.138)
Contiguity	(omitted)	–1.467
		(0.141)**
Number of Riparian States	0.075	0.095
	(0.064)	(0.015)***
Upstream Power	10.441	1.203
	(4.151)**	(2.390)
Downstream Power	–3.019	–1.588
	(5.211)	(2.868)
Years Variable	–1.330	–1.237
	(0.582)**	(0.139)***
Spline 1	–0.902	2.249
	(0.430)**	(0.532)***
Spline 2	1.113	–3.328
	(0.544)**	(0.779)***
Spline 3	(omitted)	(omitted)
Constant	–3.954	–0.395
	(0.670)***	(0.235)*
Observations	2265	2820
Log Pseudolikelihood	–96.66	–919.39
Wald χ^2	38.22	424.74
Prob > χ^2	0.00	0.00

Standard errors clustered on basin–dyads in parentheses.
* significant at 10%.
* significant at 5%.
*** significant at 1%.

Table 4
In–sample predictions.

Quintile of predicted values	Conflict (Model 1)	Cooperation (Model 2)
Least–Likely Category (1–2)	1	37
Most–Likely Category (3–5)	20	431
Total	21	468

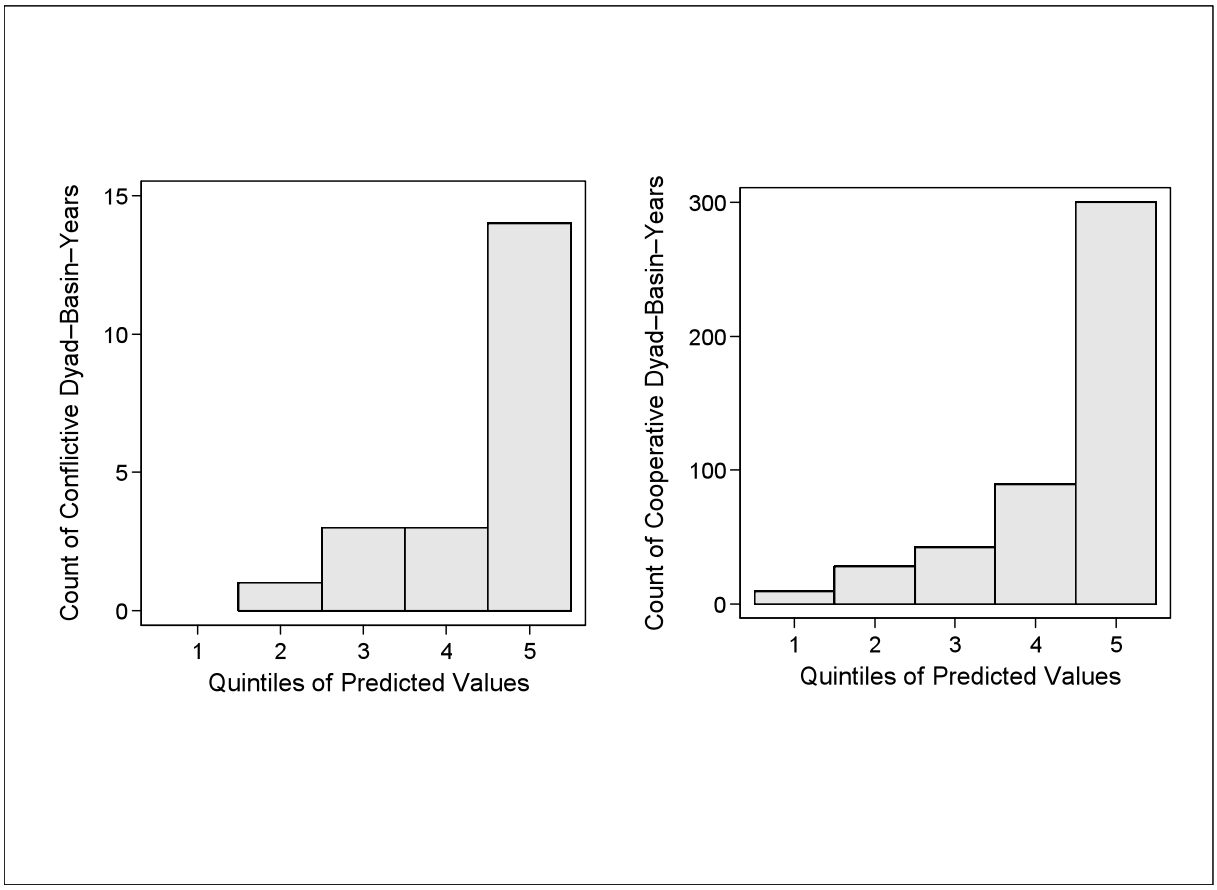

Fig. 1. In–sample predictions.

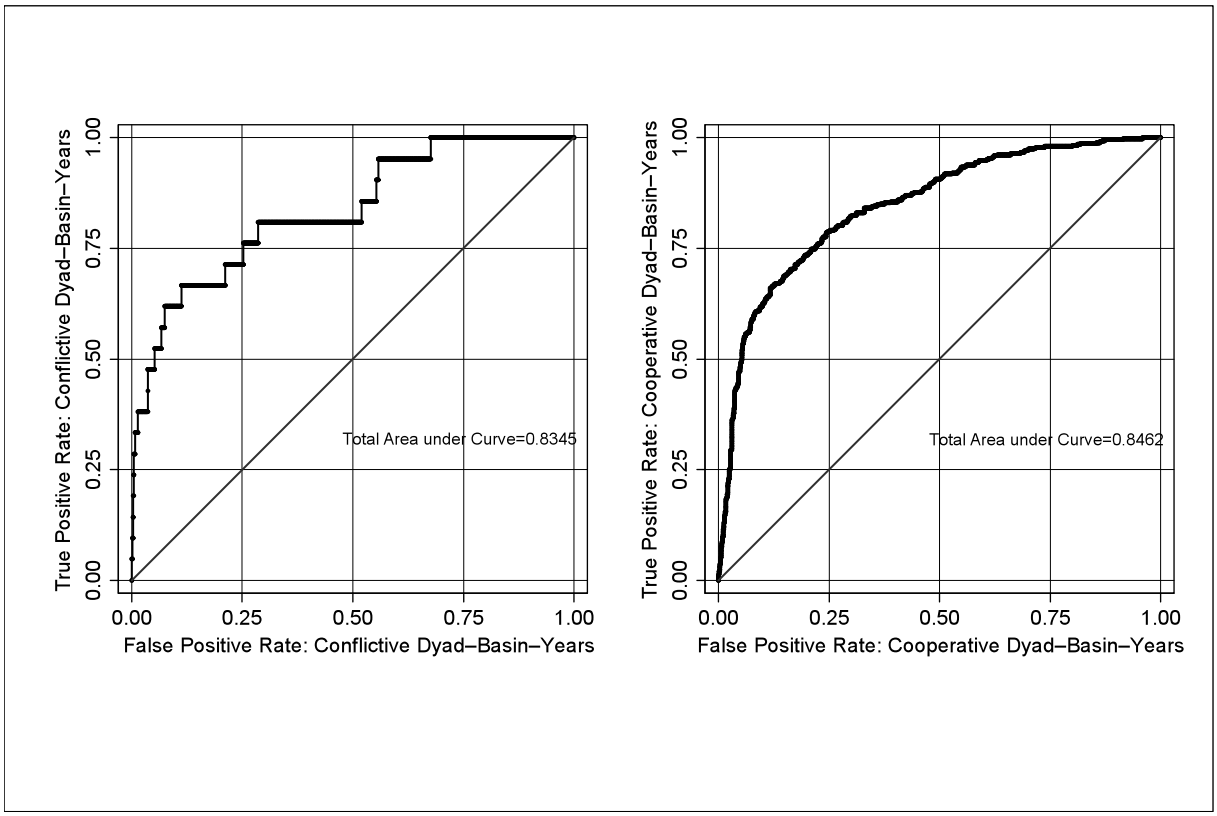

Fig. 2. In–sample predictions: ROC plots.

Table 5
Out–of–sample predictions: (Quasi–) forecasting

Quintile of predicted values	Conflict	Cooperation
Least–Likely Category (1–2)	14	34
Most–Likely Category (3–5)	20	821
Total	34	855

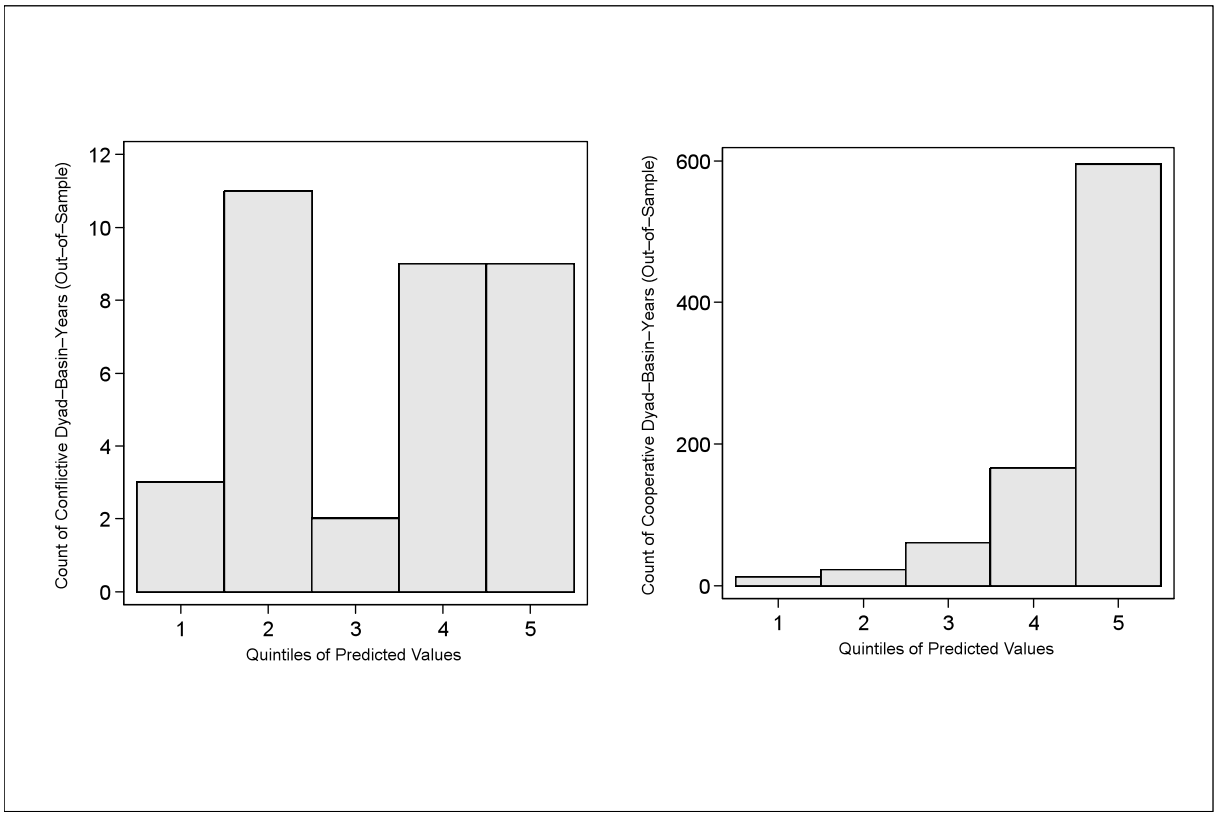

Fig. 3. Out–of–sample predictions: (Quasi–) forecasting.

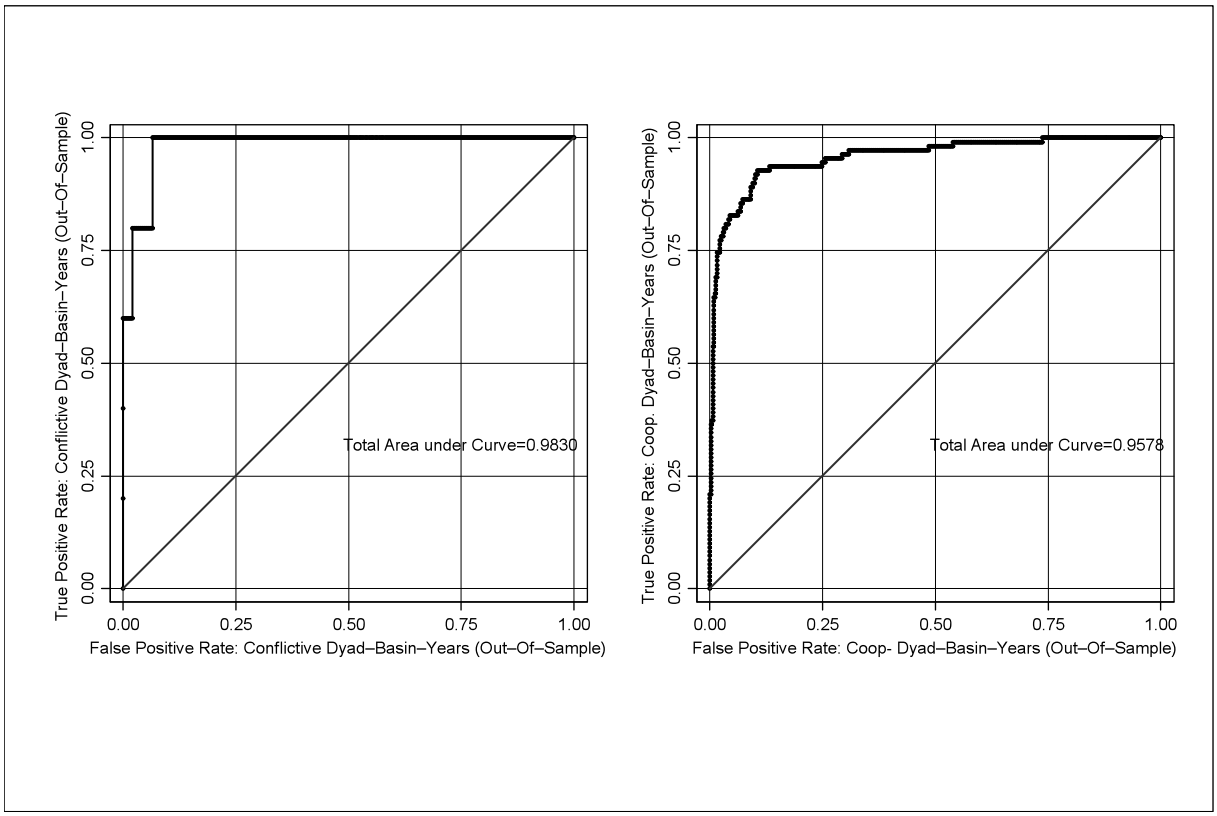

Fig. 4. Out–of–sample predictions: (Quasi–) forecasting via ROC plots.

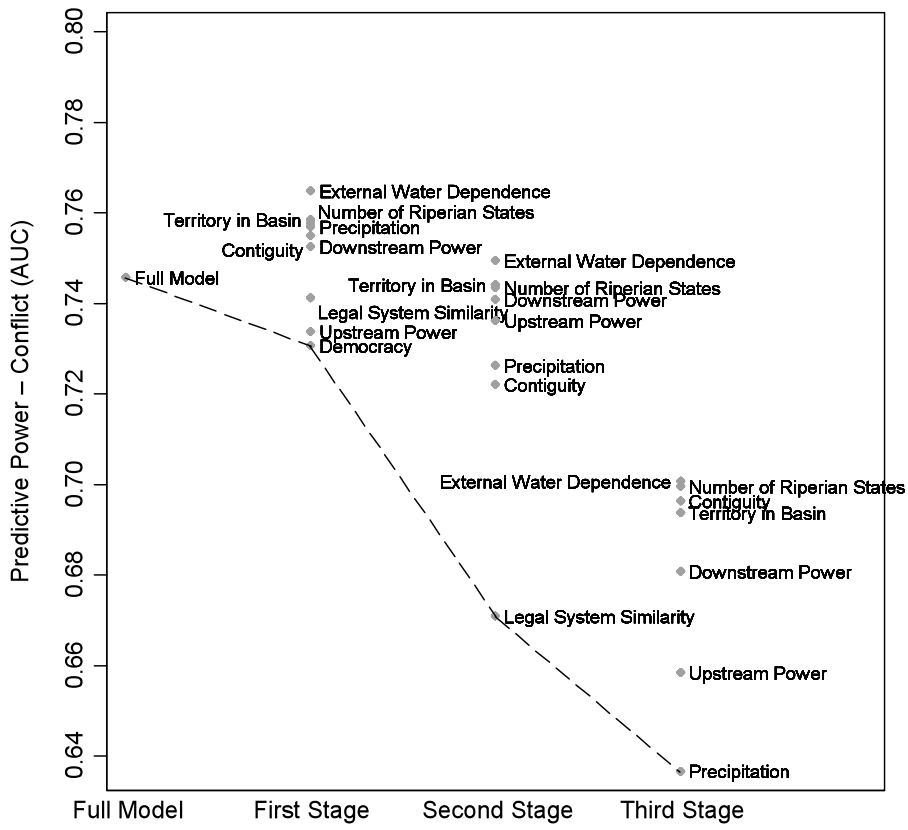

Fig. 5. Out–of–sample predictive power: International river basin conflict.

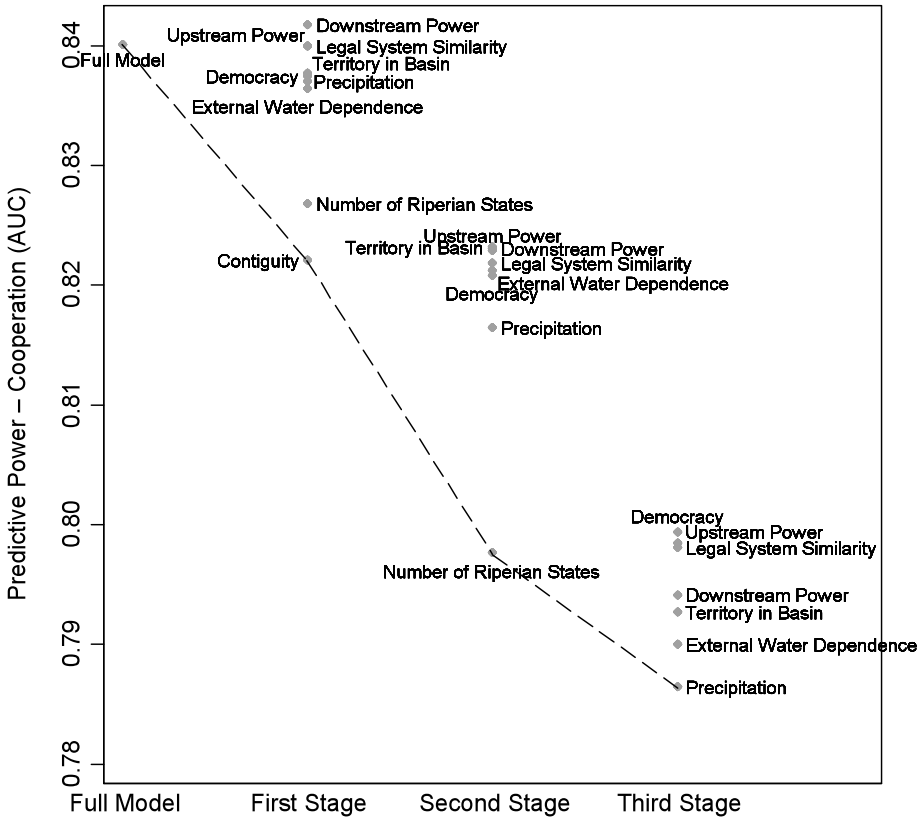

Fig. 6. Out–of–sample predictive power: International river basin cooperation.

Table 6
Identified basins at risk and cooperative basins in comparison.

Basins at risk – Yoffe et al. (2003)	Basins at risk – This paper	Cooperative basins that differ from first column
Aral Sea	Amazon	Aral Sea
Asi/Orontes	Amur	Ganges–Brahmaputra–Meghna
Ca	Aral Sea	Indus
Chiloango	Atrak	Jordan
Cross	Colorado	Kura–Araks
Drin	Danube	La Plata
Ganges–Brahmaputra–Meghna	Ganges–Brahmaputra–Meghna	Lake Chad
Han	Han	Limpopo
Indus	Helmand	Mekong
Irrawaddy	Ili/Kunes He	Nile
Jordan	Indus	Ob
Kune	Kunene	Okavango
Kura–Araks	Lake Chad	Senagal
La Plata	Niger	Tigris–Euphrates
Lake Chad	Nile	Yalu
Lempa	Ob	Zambezi
Limpopo	Pu Lun T'o	
Mekong	Samur	
Nile	Senegal	
Ob	Sujfun	
Okavango	Tigris–Euphrates	
Red	Tumen	
Saigon	Vistula/Wista	
Salween	Yalu	
Senegal	Zambezi	
Song Vam Co Dong		
Tigris–Euphrates		
Yalu		
Zambezi		

Table 7
False negatives: Dyad–basin–years in which cooperation or conflict did occur although not predicted by models.

Conflictive dyad–basin–years	Cooperative dyad–basin–years
Cancoso/Lauco	Congo
Courantyne	Cross
Dniester	Douro/Duero
Elbe	Elbe
Negro	Essequibo
	Guadiana
	Maritsa
	Mono
	Neretva
	Oder/Odra
	Orange
	Pascua
	Rhine
	Salween
	San Martin
	Struma
	Tumbes–Poyango

References

Alcamo, J., Flörke, M., Märker, M., 2007. Future long–term changes in global water resources driven by socio–economic and climatic changes. *Hydrological Sciences Journal* 52, 247–275.

Bechtel, M., Leuffen, D., 2010. Forecasting European Union politics: Real–time forecasts in political science time series analysis. *European Union Politics* 11, 309–327.

Beck, L., Bernauer, T., 2011. How will combined changes in water demand and climate affect water availability in the Zambezi river basin? *Global Environmental Change* 21, 1061–1072.

Beck, N., Katz, J., Tucker, R., 1998. Taking time seriously in binary time–series cross–section analysis. *American Journal of Political Science* 42, 1260–1288.

Beck, N., King, G., Zeng, L., 2000. The problem with quantitative studies of conflict: A conjecture. *American Journal of Political Science* 94, 21–36.

Bernauer, T., Kalbhenn, A., 2010. The politics of international freshwater resources. In: Denemark, R. (Eds.), *The international studies encyclopedia*. Wiley–Blackwell, Hoboken, NJ.

Bernauer, T., Siegfried, T., 2012. Climate change and international water conflict in Central Asia. *Journal of Peace Research* 49, 227–239.

Brochmann, M., 2012. Signing river treaties – Does it improve river cooperation? *International Interactions* 38, 141–163.

Brochmann, M., Gleditsch, N. G., 2006. *Shared rivers and international cooperation*. Unpublished manuscript.

Brochmann, M., Hensel, P., 2009. Peaceful management of international river claims. *International Negotiation* 14, 393–418.

Bueno de Mesquita, B., 2011. A new model for predicting policy choices: Preliminary tests. *Conflict Management and Peace Science* 28, 65–87.

Choucri, N., Robinson, T., 1978. *Forecasting in international relations: Theory, methods, problems, prospects*. W. H. Freeman, San Francisco, CA.

Conca, K., Wu, F., Mei, C., 2006. Global regime formation or complex institution building? The principled content of international river agreements. *International Studies Quarterly* 50, 263–285.

De Stefano, L., Edwards, P., de Silva, L., Wolf, A., 2010. Tracking cooperation and conflict in international basins: Historic and recent trends. *Water Policy* 12, 871–884.

Dinar, A., Blankespoor, B., Dinar, S., Kurukulasuriya, P., 2010. The impact of water supply variability on treaty cooperation between international bilateral river basin riparian states. *World Bank Policy Research Working Paper Series 5307*.

Dinar, A., Dinar, S., 2003. Recent developments in the literature on conflict and cooperation in international shared water. *Natural Resources Journal* 43, 1217–1287.

Espey, M., Towfique, B., 2004. International bilateral water treaty formation. *Water Resources Research* 40, W05S05 (1–8).

Furlong, K., Gleditsch, N. G., Hegre, H., 2006. Geographic opportunity and neomalthusian willingness: Boundaries, shared rivers, and conflict. *International Interactions* 32, 79–108.

Gerlak, A., Grant, K., 2009. The correlates of cooperative institutions for international rivers. In: Volgy, T., Šabič, Z., Roter, P., Gerlak, A. (Eds.), *Mapping the new world order.* Wiley–Blackwell, Hoboken, NJ. (114–147).

Gleditsch, K. S., Ward, M. D., 2013. Forecasting is difficult, especially the future: Using contentious issues to forecast interstate disputes. *Journal of Peace Research*, in press.

Gleditsch, N. G., Furlong, K., Hegre, H., Lacina, B., Owen, T., 2006. Conflicts over shared rivers: Resource scarcity or fuzzy boundaries? *Political Geography* 25, 361–382.

Goertz, G., 2006. *Social science concepts: A user's guide.* Princeton University Press, Princeton, NJ.

Goldstone, J., Bates, R., Epstein, D., Gurr, T., Lustik, M., Marshall, M., Ulfelder, J., Woodward, M., 2010. A global model for forecasting political instability. *American Journal of Political Science* 54, 190–208.

Hensel, P., Mitchell, S., Sowers, T., Thyne, C., 2008. Bones of contention: Comparing territorial, maritime, and river issues. *Journal of Conflict Resolution* 52, 117–143.

Kalbhenn, A., 2011. Liberal peace and shared resources – A fair–weather phenomenon? *Journal of Peace Research* 48, 715–735.

Kalbhenn, A., Bernauer, T., 2012. *International Water Cooperation and Conflict: A New Event Dataset*. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2176609.

Marshall, M., Jaggers, K., 2004. *POLITY IV project: Political regime characteristics and transitions*. *Dataset user's manual*. University of Maryland, College Park, MD.

O'Brian, S., 2010. Crisis early warning and decision support: Contemporary approaches and thoughts on future research. *International Studies Review* 12, 87–104.

Owen, T., Furlong, K., Gleditsch, N. G., 2004. *Codebook for the shared river basin and GIS database*. Peace Research Institute, Oslo.

Schneider, G., Gleditsch, N. P., Carey, S., 2010. Exploring the past, anticipating the future. *International Studies Review* 12, 1–7.

Schneider, G., Gleditsch, N. P., Carey, S., 2011. Forecasting in international relations: One quest, three approaches. *Conflict Management and Peace Science* 28, 5–14.

Singer, D., 1988. Reconstructing the Correlates of War dataset on material capabilities of states, 1816–1985. *International Interactions* 14, 115–132.

Singer, D., Bremer, S., Stuckey, S., 1972. Capability distribution, uncertainty, and major power war, 1820–1965. In: Russett, B. (Ed.), Peace, war, and numbers. SAGE, Beverly Hills, CA (19–48).

Singer, D., Wallace, M., 1979. *To augur well: Early warning indicators in world politics*. SAGE, Beverly Hills, CA.

Stinnett, D., Tir, J., 2009. The institutionalization of river treaties. *International Negotiation* 14, 229–251.

Tir, J., Ackerman, J., 2009. Politics of formalized river cooperation. *Journal of Peace Research* 46, 623–640.

UCDP., 2008. *UCDP database*. Department of Peace and Conflict Research, Uppsala University, Uppsala.

Vörösmarty, C., McIntyre, P., Gessner, M., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S., Sullivan, C., Reidy Liermann, C., Davies, P., 2010. Global threats to human water security and river biodiversity. *Nature* 467, 555–561.

Ward, M. D., Greenhill, B. D., Bakke, K. M., 2010. The perils of policy by p–value: Predicting civil conflicts. *Journal of Peace Research* 47, 363–375.

Ward, M. D., Siverson, R. M., Cao, X., 2007. Disputes, democracies, and dependencies: A reexamination of the Kantian peace. *American Journal of Political Science* 51, 583–601.

Wolf, A., Natharius, J., Danielson, J., Ward, B., Pender, J., 1999. International river basins of the world. *Water Resources Development* 15, 387–427.

Wolf, A., Stahl, K., Macomber, M., 2003. Conflict and cooperation within international river basins: The importance of institutional capacity. *Water Resources Update* 125, 1–10.

Wolf, A., Yoffe, S., Giordano, M., 2003. International waters: Identifying basins at risk. *Water Policy* 5, 29–60.

Yoffe, S., Wolf, A., Giordano, M., 2003. Conflict and cooperation over international freshwater resources: Indicators of basins at risk. *Journal of the American Water Resources Association* 39, 1109–1125.

Yoffe, S., Fiske, G., Giordano, M., Larson, K., Stahl, K., Wolf, A., 2004. Geography of international water conflict and cooperation: Data sets and applications. *Water Resources Research* 40, 1–12.

Zawahri, N., Gerlak, A., 2009. Navigating international river disputes to avert conflict. *International Negotiation* 14, 211–227.

Zawahri, N., Mitchell, S., 2011. Fragmented governance of international rivers: Negotiating bilateral vs. multilateral treaties. *International Studies Quarterly* 55, 1–24.

Zeitoun, M., Mirumachi, N., 2008. Transboundary water interaction I: Reconsidering conflict and cooperation. *International Environmental Agreements* 8, 297–316

Zeitoun, M., Mirumachi, N., Warner, J., 2010. Transboundary water interaction II: Soft power underlying conflict and cooperation. *International Environmental Agreements* 11, 159–178