Supplementary materials

Title: Has the Three-Gorges Dam made the Poyang Lake wetlands wetter and drier?

Data used in the analyses

The first dataset consists of daily rainfall (mm) from 20 stations in the catchment area upstream of the Three-Georges Dam (3GD), daily water surface evaporation rates (mm) from three stations near the dam, and river discharge (m^3/s) immediately downstream of the dam for the period 1968-2008. All together (excluding 2003), there are 12784 daily records before 2003 and 1827 daily records after.

The second dataset consists of daily water levels (m) of the Poyang Lake from four stations, daily rainfall (mm) from 12 stations in the lake's catchment area and daily discharges from five local rivers to the lake for the period 1968-2008. Some lake water level data were missing: 1-6 September and 29 September-21 October 1988, and 15-31 August and 1-13 October in 1989.

Forming informative explanatory variables

A binary indicator variable (g) was formed, with g = 0 for the period 1968-2002 and g = 1 for 2003-2008. This variable g in regression models indicated the categorical effect associated with absence and presence of the 3GD during the two periods, respectively. Another indicator variable BA was also formed to indicate before and after day 250 of each year: i.e., each year was divided into two periods corresponding with the dam operation (P1 and P2). The first 250 d (P1) covered the period including the wet season that finishes in early September. Special attention was paid to 2003 data as then a large amount of water was stored for the newly constructed dam.

On average, the annual discharge was 404 billion m^3 prior to the dam construction and 386 billion m^3 (BCM) afterwards. For the same rainfall condition, 4.4% less water discharges downstream of the dam compared with the condition prior to the dam.

The average rainfalls (mm/day) over the whole period (1968-2008, excluding 2003) were 3.18 and 1.98 (mm/day) for days before (P1) and after (P2) day 250. The average discharge rate after September (> 250 d) for the 2004-2008 period was 528.33 BCM/m-rain, which is considerably smaller than that in the same period for all the previous 5-year averages over the 25 years prior to the dam construction (shown in Table 1 of the paper with the details given in Table S1 on a monthly basis). For the days before September, the discharge rate is similar for both pre- and post-dam years. The average ratios of after (g = 1) and before (g = 0) the 3GD for an effective catchment area (flow/rainfall) are 0.973 and 0.933 for P1 and P2, respectively.

Table S1. Comparison of upstream rainfall, river discharge at the dam and their ratios for the two periods (subscript 1 for before 3GD and 2 for after 3GD) based on monthly averages.

Month	$rain_1$	$rain_2$	ratio	$discharge_1$	$discharge_2$	ratio	$discharge_1/rain_1$	$discharge_2/rain_2$	ratio
1	0.51	0.51	0.99	368.55	402.57	1.09	723.25	797.01	1.10
2	0.63	0.93	1.47	335.23	393.83	1.17	533.74	425.42	0.80
3	1.13	1.29	1.15	374.37	470.24	1.26	332.36	363.79	1.09
4	2.71	2.60	0.96	585.06	629.34	1.08	216.16	242.42	1.12
5	4.31	4.21	0.98	995.94	961.16	0.97	231.06	228.13	0.99
6	5.58	4.58	0.82	1602.35	1478.38	0.92	287.05	322.91	1.12
7	6.12	6.27	1.02	2553.81	2164.04	0.85	417.52	345.41	0.83
8	4.90	4.43	0.90	2264.71	2030.92	0.90	462.24	458.66	0.99
9	4.43	4.20	0.95	2117.38	1937.27	0.91	477.76	461.76	0.97
10	2.70	2.77	1.03	1528.15	1163.77	0.76	565.08	419.57	0.74
11	1.25	1.29	1.03	837.13	827.81	0.99	668.47	642.87	0.96
12	0.52	0.41	0.78	498.17	484.44	0.97	952.09	1193.96	1.25

Statistical Modeling

To model nonlinear relationships between the responses of hydrological variables (discharge at the dam and PY daily water level changes in our case) to forcing conditions, we adopted the semi-parametric approach of generalized additive models (GAMs) to estimate effects of each individual covariate non-parametrically. A generalized additive model (GAM) is a generalized linear model (GLM) in which the linear predictor is given by a sum of smooth functions of the covariates and a conventional parametric component of the linear predictor. By allowing non-parametric fits, well-designed GAMs provide a good match of the training data with relaxed assumptions on the actual relationship. The link function S_g relates the expected value of the distribution to the predictors (co-variates).

Generalized additive models [3] have been extensively used in environmental, biological and medical studies [2]. The **R** [3] package mgcv[1] makes this approach widely available. Hydrological phenomena can be represented as either linear or spline terms. If a linear function approximates the relationship well, it is preferred to the (more complex) spline term. More details of the method can be found in [3]~~[3~~,[4]~~4]~~. In our case, we replaced some spline terms (which appear to be linear) with linear terms and found little change in the results. Also, we used an identity link with the mean discharge given by: E (Flow) = $X\beta + s_g$ (d), where X is the design matrix, and s_g are two smooth cyclic functions for 1978-2002 (g = 0) and for 2004-2008 (g = 1). These functions capture changes in the discharge due to unknown sources (those not expressed in the model with specific data input), including evaporation and possible underground leakage. Seasonal effects due to these sources are embedded in the functions.

(1) <u>*Analysis of the river discharge immediately downstream of the dam*</u>

The analysis and models used, as input variables, daily upstream rainfall data up to previous 14 d (Rain.i, i = 1,2, … 14) and averaged rainfall data aggregated for 15-21 d (p15-21), 22-35 d (p.22-35) and 36-63 d (p.36-63). Rainfalls beyond 63 d were found not significant and hence not included. The predictive variables used were:

rain: rainfall on the given day

[1] http://cran.r-project.org/web/packages/mgcv/index.html, last accessed 17 July 2012.

w.rain.1: rainfall on the previous day

w.rain.2: rainfall from two days before

arain3-5: average rainfall from the previous 3-5 d

arain6-8: average rainfall from the previous 6-8 d

arain9-11: average rainfall from previous 9-11 d

arain12-14: average rainfall from previous 12-14 d

p.rain.15-21: average rainfall from previous 15-21 d

p.rain.22-35: average rainfall from previous 22-35 d

p.rain.36-63: average rainfall from previous 36-63 d

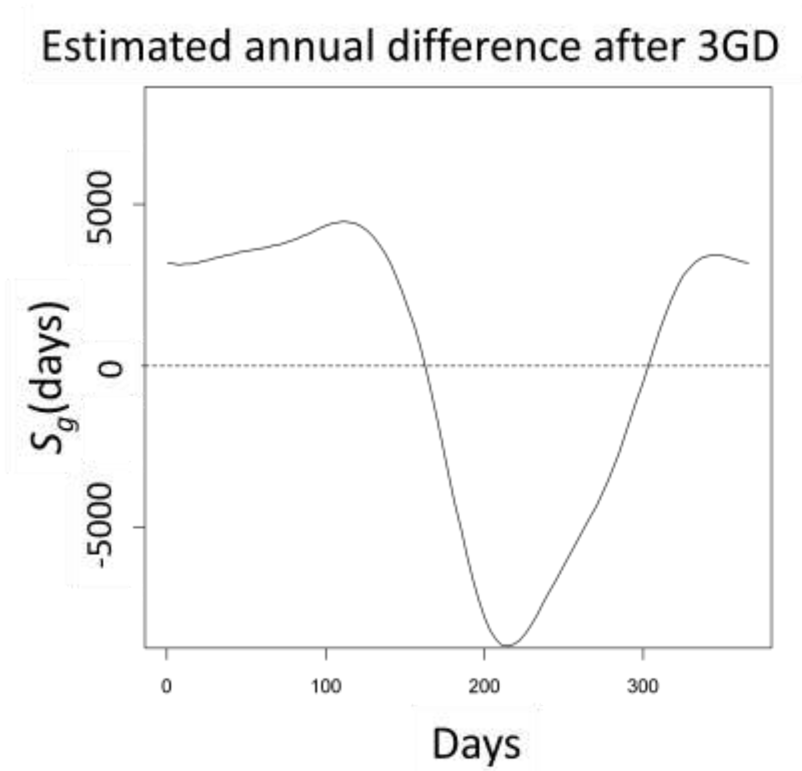

Fig. S1 Cyclic function.

The design matrix *X* in the semi-parametric regression models included all the main effects and up to three way interactions of all the rainfall variables with indicator variables *g* and BA. Two cyclic smooth functions were used to estimate the discharge pattern for *g* = 0 (before 2003) and 1 (after 2003). The adjusted R-square for the regression was 0.86.

The least-squares method was used for parameter estimation. Data was weighted equally, instead of weighting based on the inverse of the variance [1]~~[e.g., 1]~~. In our case, weighting was undesirable because the weighted error led to poor fits. The residuals are assumed to follow a first-order autoregressive process. For computational convenience, this error process was imposed within each year. The p-values for the fixed terms in the GAM model based on the modeling results were small, indicating the statistical significance of the estimates of these terms. The difference between the years before and after the dam, estimated by the smooth function, was highly significant (p-value < 0.0001). The cyclic function is plotted in Fig. S1.

Prediction

The established models were then applied to make predictions based on a variety of annual rain patterns: each individual year rain pattern, and three different rain patterns averaged over 1968-2002, 2004-2008 and all years.

(2) *<u>PY Lake water level analysis</u>*

The variables of rainfall upstream of the 3GD, as defined above in the river discharge models, were used in the lake water level models, i.e., *arain, w.rain.1, w.rain.2: arain3-5arain6-8, arain9-11, arain12-14, p.rain.15-21, p.rain.22-35, p.rain.36-63*. Local rainfall data in the Poyang lake catchment area were also included: daily rainfall for previous 12 d, *py.rain.i* (i = 1, 2, …, 12). Rainfalls earlier were found to play no role and hence were not further modeled. Again, two smooth cyclic functions for 1978-2002 (g = 0) and for 2004-2008 (g = 1) were used to capture changes in the lake water level due to unknown sources, including particularly evaporation.

The design matrix X in the semi-parametric regression models included all the main effects and up to three-way interactions of all the upstream rainfall variables with indicator variables g and BA together with all the 12 local rain fall variables (*py.rain.i*) and their two-way interactions with g. Again, two cyclic smooth functions were used to estimate the water level for g = 0 (before 2003) and 1 (after 2003).

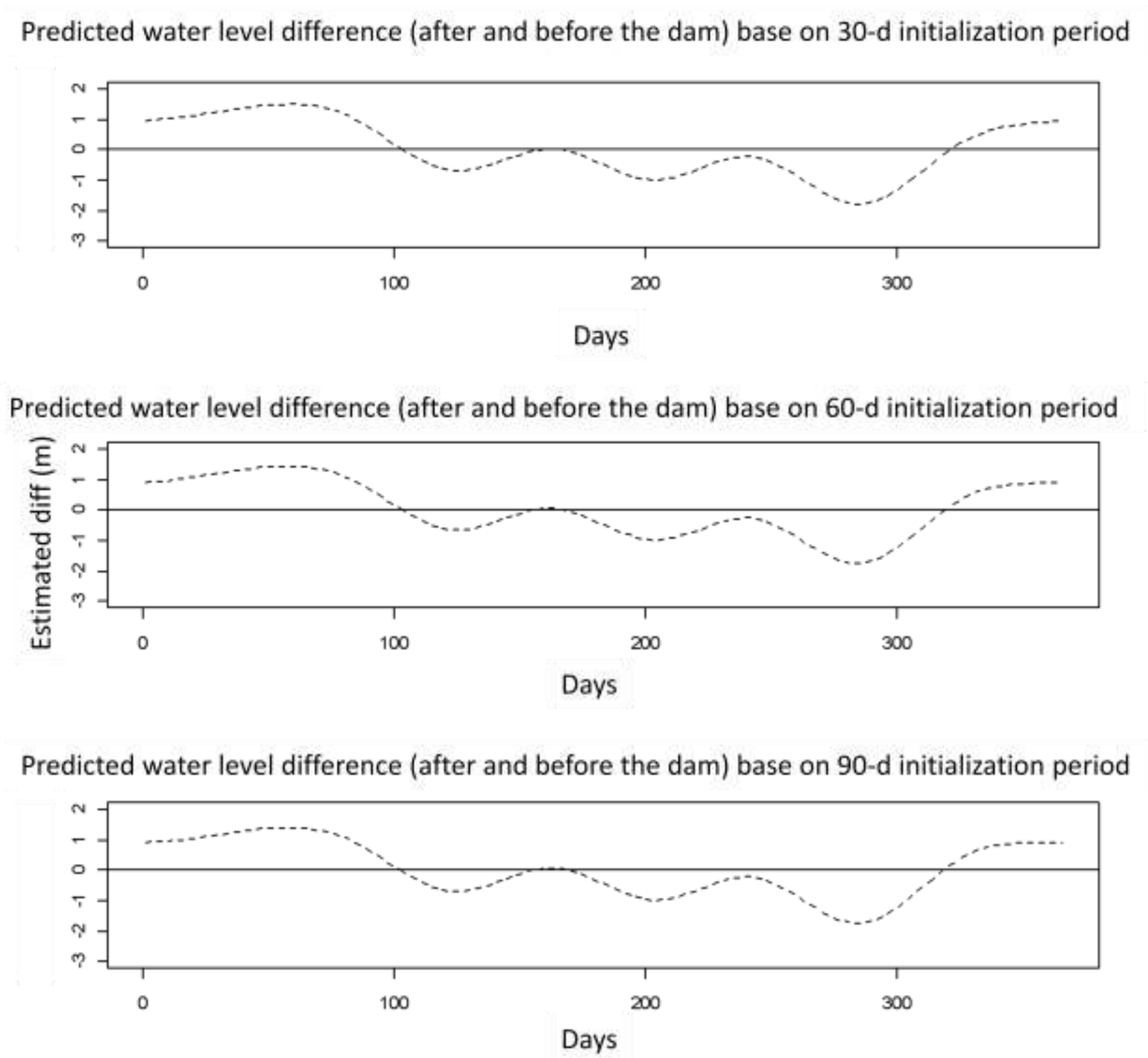

Fig. S2 Predicted lake water level differences.

The water level prediction depends on water levels on previous days (i.e., how much water was already stored in the lake). It was therefore necessary to set the model conditional on the water level at priori. In the analysis, the model was set to be conditional on the water level 30 d, 60 d and 90 d earlier. These three test cases produced very similar results (Fig. S2). The modeling results show that the difference between the years before and after the 3GD is very significant (p-value < 0.0001), again with small p-values for all the fixed terms in the GAM model.

None of the interaction terms between rainfall in the PY lake area and *g*, or rainfall in the 3GD upstream catchment and *g* was found to be significant. This indicates that the relationships between the water level and the rainfall (local and upstream) were not changed after 3GD. However, there are significant interactions of upstream rainfall with BA and *g*, which indicates some control effect of 3GD due to storing water before September and releasing water after.

The smoothed temporal (daily) effect was extracted with the difference in the water level change between the pre- and post-dam periods as shown below.

References

[1] Barry, D. A., 1990. Comments on "Estimating Michaelis-Menten or Langmuir isotherm constants by weighted nonlinear least squares" by P. Persoff and J. F. Thomas. *Soil Science Society of America Journal*. 54: 941-942.

[2] Hastie, T., Tibshirani, R., 1986. Generalized additive models. *Statistical Science*, 1: 297-310.

[3] Wood, S. N., 2006. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC Press.

[4] Wood, S. N., 2008. Fast stable direct fitting and smoothness selection for generalized additive models. *Journal of the Royal Statistical Society* (B), 70: 495-518.