

# Measuring Predictability of Daily Streamflow Processes Based on Univariate Time Series Model

Wen Wang<sup>a</sup>, Pieter H.A.J.M. Van Gelder<sup>b</sup>, and J. K. Vrijling<sup>b</sup>

<sup>a</sup> *State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, 210098, China*

<sup>b</sup> *Faculty of Civil Engineering & Geosciences, Section of Hydraulic Engineering, Delft University of Technology, 2628 CN Delft, Netherlands*

**Abstract:** Predictability is an important aspect of the dynamics of hydrological processes. The predictability of streamflow processes can be estimated based on the multivariate approach, which takes multiple explanatory variables into consideration, or, based on a univariate time series approach, which measures the predictability on the basis of univariate streamflow itself. In this study we investigate the predictability of 31 daily average discharge series with different drainage areas observed in 8 river basins in Europe and northern America on the basis of univariate time series approach. The results show that, although the existence of long memory is detected in the daily streamflow processes, the predictability of the streamflow process is more dominated by short-range autocorrelations than by the existence of long-memory in the streamflow process; the predictability is positively related to watershed scale, that is, the larger the watershed scale, the better the predictability of the streamflow process, and this kind of relationship mainly stems from the positive relationship between autocorrelation and basin scale. Because of the impacts of many factors, the predictability is dynamic rather than invariable, and there are many uncertainties present in the estimation of streamflow predictability.

**Keywords:** predictability; time series; long memory; ARFIMA model; AR model.

## 1. INTRODUCTION

Predictability refers to the ability to make predictions of future events based on either past information or a theoretically complete knowledge of the physical system. The predictability of streamflow processes can be estimated based on the multivariate approach, which takes multiple explanatory variables into consideration, or, based on a univariate time series based approach, which measures the predictability on the basis of univariate streamflow itself. The efforts in evaluating predictability of real-world processes so far are in fact concentrating on improving long-term forecasting accuracy by finding better predictors, which is essentially a multivariate approach. Although making forecasts is normally an inevitable step in the procedure of assessing predictability, the issue of forecast accuracy should not be confused with the issue of predictability, and efforts to improve forecasts should not be confused with approaches to understanding the predictability, because predictability is a physical system property that depends on intrinsic dynamics. It is believed [Bloschl and Zehe, 2005] that to learn how to separate the predictable and the unpredictable would be an exciting research field in hydrology in the coming years. It would be also interesting to investigate how each factor, such as El Niño-Southern Oscillation (ENSO) and the Arctic Oscillation (AO) [Hastenrath, 1990; Maurer and Lettenmaier, 2003; Berg and Mulroy, 2006], contributes to the overall predictability of streamflow processes. On the other hand, because river runoff gives an integral measure of

the hydrometeorological conditions in a catchment as a whole, it would therefore be interesting to investigate how the signal of previous streamflow, a basic predictable component of hydrologic systems at watershed scale, contributes to streamflow predictability. However, that is an issue not well recognized.

In addition, the evidences of the presence of long memory property has been revealed in many observed hydrological time series [e.g., Mandelbrot and Wallis, 1969; Hosking, 1984; Wang et al., 2007; Mudelsee, 2007]. The calculations based on comparing the innovation variance and unconditional variance of stationary series suggest that long-memory fractionally integrated autoregressive moving average (ARFIMA) processes often have quite long predictable memory and that fractional integration extends the prediction memory of an ARMA (integrated autoregressive moving average) process [Andersson, 2000]. Hence, a special concern in the present study is that whether we can see long predictability in an observed hydrological series even though we can detect the existence of long-memory.

Because the streamflow flow series is often treated as a univariate time series and can be modelled conventionally with ARMA type models, we therefore take a univariate time series based approach [Wang, et al., 2004] to measure the predictability of univariate streamflow time series in the present study. The paper is organized as follows. In Section 2, the theoretical predictability of ARMA process and ARFIMA process will be analyzed as a basis for comparing with the univariate streamflow processes. In Section 3, the predictability of 31 univariate streamflow series at various sites over the world will be analyzed. Some discussions will be given in Section 4 and the results are concluded in Section 5.

## 2. MEASURING PREDICTABILITY OF A TIME SERIES

### 2.1 Definition of time series predictability

Clements and Hendry [1998] define a random variable  $x_t$  to be unpredictable with respect to an information set  $\Omega_{t-1}$  if the conditional distribution  $F_{x_t}(x_t | \Omega_{t-1})$  and the unconditional distribution  $F_{x_t}(x_t)$  of  $x_t$  coincide, i.e. if

$$F_{x_t}(x_t | \Omega_{t-1}) = F_{x_t}(x_t). \quad (1)$$

This notion of unpredictability implies that the information set  $\Omega_{t-1}$  does not improve the prediction of  $x_t$ . If  $\Omega_{t-1}$  is restricted to past realizations of  $x_t$ , then (1) implies that past realizations do not help to predict  $x_t$ .

A weaker condition would only require that the conditional variance of the residual series  $\{e_t\}$  equals the constant unconditional variance  $\sigma_x^2$ , i.e.,

$$\text{Var}(e_t | \Omega_{t-1}) = \text{Var}(e_t) = \sigma_x^2. \quad (2)$$

for all  $t$ . The mean value as well as the volatility of a time series is said to be predictable if (2) does not hold [Raunig, 2006].

### 2.2 Theoretical predictability for ARMA process or ARFIMA process with known formula

To see if random variable  $x_t$  is predictable, we need to know the conditional variance  $\text{Var}(e_t | \Omega_{t-1})$ , or simply  $\text{Var}(e_t)$ , in (2), and the constant unconditional variance  $\sigma_x^2$ . Theoretical formulae are available for calculating both of them for the ARFIMA model and its reduced version ARMA model. Consequently, we refer to the predictability obtained in this way as the theoretical predictability.

The general form of ARFIMA( $p, d, q$ ) model is given by

$$(1 - \phi_1 B - L \phi_p B^p)(1 - B)^d x_t = (1 + \theta_1 B + L \theta_q B^q) a_t \quad (3)$$

where  $|d| < 0.5$ ,  $B$  is the back shift operator, i.e.,  $Bx_t = x_{t-1}$  and  $a_t$  is i.i.d. with mean zero and variance  $\sigma_a^2$ . When  $d = 0$ , the ARFIMA model is reduced to an ARMA( $p, q$ ) model. If  $q = 0$ , the ARMA( $p, q$ ) model if further reduced to an AR( $p$ ) model.

According to the Wold decomposition theorem, under stationarity, the process variance  $\sigma_x^2$  and  $h$ -step ahead optimal forecast error variance  $\xi^2(h)$  of (3) are given as:

$$\begin{cases} \sigma_x^2 = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j^2, \psi_0 = 1 \\ \xi(h) = \sigma_a^2 \sum_{j=0}^{h-1} \psi_j^2, \psi_0 = 1 \end{cases} \quad (4)$$

where  $\psi_j$  are given by:

$$(1 + \psi_1 B + \psi_2 B^2 + L) = \frac{(1 + \theta_1 B + L \theta_q B^q)}{(1 - \phi_1 B - L \phi_p B^p)(1 - B)^d}$$

Granger and Newbold [1986, p. 310] proposed a measure of predictability for covariance stationary series, as the ratio of the variance of the forecast error to the variance of the original time series. According to the definition of Granger and Newbold [1986], predictability is given by  $R_{h|x}^2 = 1 - \xi(h) / \sigma_x^2$ . With the increase of  $h$ ,  $\xi(h)$  and  $\sigma_x^2$  will get close. At a certain lead time  $H$ , they get sufficiently close so that  $\xi(H) / \sigma_x^2 < c$ , where  $c$  is a constant close to 1 (say, 0.95), whereas for times larger than  $H$  (i.e.,  $H+$ ),  $\xi(H+) / \sigma_x^2 \geq c$ , i.e.,  $R_{H+|x}^2 \leq 1 - c$ . Therefore, given a  $c$ , we could get a  $H$ , which satisfies

$$\xi(H) < c\sigma_x^2 \leq \xi(H+1). \quad (5)$$

The estimate given by (5) means that the model forecasts are not more accurate than the mean value of the process after  $H$  steps at a given level  $1-c$ . Consequently, instead of using the general definition in (1) or (2), we can define the predictability of a stationary process more precisely as the predictable horizon  $H$  after which the prediction is no better than the mean value. Because the predictability defined in this way is based on theoretical formula, we refer it to as theoretical predictability (TP).

### 2.3 Sample predictability for a process with unknown formula

As mentioned above, TP is calculated based on the known model of covariance stationary series, for which the formula are available for calculating the variance of the process and the variance of multi-step forecast error. However, the true model of a given time series is rarely known, especially for real-world processes. Alternatively, we may take a forecast error based approach, which measures the predictability based on forecast errors. Correspondingly, such type of predictability is referred to as sample predictability (SP), because it is estimated from forecast error samples.

To measure the SP, we may use the coefficient of efficiency ( $CE$ ) proposed by Nash and Sutcliffe [1970] which is widely adopted as a model performance measure in the hydrology community, given by

$$CE = 1 - \frac{\sum_{i=1}^n (Q_i - \hat{Q}_i)^2}{\sum_{i=1}^n (Q_i - \bar{Q})^2}, \quad (6)$$

where  $n$  is the data size,  $Q_i$  is the observed value,  $\hat{Q}_i$  is the predicted value,  $\bar{Q}$  is the mean value of the observed data. When measuring predictability for a observed processes, we should split the entire data into two parts: one calibration data set which is used to build a model, and another validation data set (or out-of-sample data) that is used for calculating  $CE$  at different lead times. Notice that, in calculating  $CE$  for measuring predictability, the mean value  $\bar{Q}$  should be calculated on the basis of calibration dataset rather than validation dataset, because by definition, the predictability is the degree that the past can be used to predict the future. This is what differs between measuring predictability from evaluating hydrological model performance, in which  $\bar{Q}$  is calculated based on the validation dataset.

In fact, the predictability measure  $R_{h|x}^2$  proposed by Granger and Newbold [1986] is essentially the same as  $CE$ . The difference between  $R_{h|x}^2$  and  $CE$  is that the former is

calculated based on theoretical formula whereas the latter is calculated based on forecast errors for sample data. With  $CE$ , the predictability of a stationary process is re-defined as the predictable horizon  $H$  after which the prediction is no better than the mean value for the process at a given level  $CE = CE_H$ , where  $CE_H$  is a small value less than 1 (say, 0.1 or 0.5), which is related to  $c$  by  $c = 1 - CE_H$ .

### 3. PREDICTABILITY OF UNIVARIATE STREAMFLOW PROCESSES

#### 3.1 Daily streamflow data used

Daily average discharge series recorded at 31 gauging stations in eight basins in Europe, Canada and USA are analyzed in the present study. The data come from Global Runoff Data Centre (GRDC) (<http://grdc.bafg.de>), US Geological Survey Water Watch (<http://water.usgs.gov/waterwatch>), and Water Survey of Canada (<http://www.wsc.ec.gc.ca>). We generally have the following three rules to select gauging stations in each basin:

- (1) The basins where the stations are located covers different geographical and climatic regions;
- (2) The drainage area of each station is basically at 5 different watershed scales, namely,  $> 10^6 \text{ km}^2$ ;  $10^6 \sim 10^5 \text{ km}^2$ ;  $10^5 \sim 10^4 \text{ km}^2$ ;  $10^4 \sim 10^3 \text{ km}^2$ ;  $< 10^3 \text{ km}^2$ ;
- (3) The stations are located in the main river channel of the river if possible. When stations at the main channel are not available, stations at major tributaries are used.

For each station, we select a segment of historical daily streamflow records of mostly 30 years long. However, because of data limitation, the shortest series covers a period of only 14 years. The segments are chosen with following criteria:

- (1) The series should be approximately stationary, as least by visual inspection. We have stationarity as our primary criterion because, when certain types of nonstationarity are present, many long-memory parameter estimators may fail.
- (2) The recording period of the data should be as early in time as possible, assuming that the influence of human intervention would be less intensive in early period (in early 20<sup>th</sup> century or even late 19<sup>th</sup> century) than in later period.
- (3) The temporal spans of streamflow series at different locations in one basin should be as close as possible, so as to mitigate possible impacts of regional low-frequency climatic variations.

The selected stations and corresponding daily streamflow series are listed in Table 1.

#### 3.2 Calculation of sample predictability of daily streamflow processes

Because  $CE$  is a measure comparing the predicted values with the overall mean value, it is not effective enough to evaluate the predictions for those series whose mean values change with seasons, which is almost always the case for hydrological processes like streamflow processes. Therefore, a seasonally-adjusted coefficient of efficiency ( $SACE$ ) [Wang et al., 2004] is used here for evaluating the predictability.  $SACE$  is calculated by

$$SACE = 1 - \frac{\sum_{i=1}^n (Q_i - \hat{Q}_i)^2}{\sum_{i=1}^n (Q_i - \bar{Q}_m)^2}, \quad (7)$$

where  $m$  is the “season”,  $m = i \text{ mod } S$  (mod is the operator calculating the remainder), ranging from 0 to  $S-1$ ; and  $S$  is the total number of “season” (Note that, a “season” here is not a real season. It may be a month or a day depending on the timescale of the time series. For daily streamflow series, one season is one day over the year.);  $\bar{Q}_m$  is the mean value of season  $m$ .

In the present study, we use  $SACE$  to measure SP of daily streamflow processes. Each daily streamflow data series are split into two parts: a calibration data set which is used to build a model, and a validation data set (or out-of-sample data) that is not used for model fitting.  $\bar{Q}_m$  for each day in (7) is calculated based on the calibration dataset. We estimate the predictability at two  $SACE$  levels 0.1 and 0.5, corresponding to  $c = 0.9$  and 0.5 in (5).

To measure SP of a process, we need a suitable forecasting model together with the model performance measure. In the present study, we use autoregressive (AR) model for all streamflow series. Each AR model is built based on the calibration dataset, then applied to the validation dataset for make predictions of different lead times.

For each out of the 31 streamflow series, the long-memory parameter (or fractional differencing parameter)  $d$  is estimated [Wang et al., 2007] with the method implemented in S-Plus. For a stationary long-memory series,  $0 < d < 0.5$ . The larger  $d$ , the longer memory the stationary process has.

The results of both the long-memory parameter estimates and predictability estimates that are represented by predictable horizon are reported in Table 1. The following facts could be revealed from Table 1:

Table 1 Selected daily streamflow time series and the estimates of their predictability

Basin	Location of gauging stations	Area (km <sup>2</sup> )	Period	ACF(1)	$d$	Predictability	
						SACE = 0.1	SACE = 0.1
Colorado	Colorado River At Lees Ferry	289,400	1922-1951	0.9738	0.4478	98	8
	Colorado River Near Cisco	62,390	1923-1952	0.9627	0.4506	18	5
	Colorado River Near Kremmling	6,167	1904-1918	0.9431	0.4863	17	5
	Williams Fork Near Parshall	476	1904-1924	0.9549	0	34	11
Columbia	Columbia River At The Dalles	613,565	1880-1909	0.991	0.4615	45	15
	Columbia River at Trail	88,100	1914-1936	0.9966	0.4187	26	13
	Columbia River at Nicholson	6,660	1933-1962	0.9778	0.4392	21	9
	Columbia River Near Fairmont Hot Springs	891	1946-1975	0.9676	0.4213	20	9
Danube	Danube river at Orsova	576232.	1901-1930	0.9931	0.2634	55	17
	Danube river at Achleiten	76653.	1901-1930	0.9577	0.3598	26	4
	Inn river at Martinsbruck	1945.	1904-1933	0.9326	0.4059	9	2
Fraser	Fraser River at Hope	217,000	1913-1942	0.9772	0.3878	32	8
	Fraser River at Shelley	32,400	1950-1979	0.9734	0.3529	12	3
	Fraser River at Mcbride	6,890	1959-1988	0.9582	0.1886	9	3
	Canoe River below Kimmel Creek	298	1972-1994	0.9294	0.3100	5	1
Mississippi	Mississippi River At Vicksburg	2,962,974	1932-1961	0.9961	0.3909	73	15
	Mississippi River at Clinton	221,608	1874-1903	0.9921	0.4001	33	13
	Minnesota River At Mankato	38,574	1943-1972	0.9917	0.4847	17	7
	Minnesota River At Ortonville	3,003	1943-1972	0.9563	0	49	10
Missouri	Missouri River at Hermann	1,353,000	1929-1958	0.9711	0.4238	28	4
	Missouri River at Bismarck,	482,776	1929-1953	0.9805	0.4124	15	2
	Missouri River at Fort Benton	64,070	1891-1920	0.9165	0	38	7
	Madison River near McAllister	5,659	1943-1972	0.9522	0	45	7
Ohio	Ohio River At Metropolis	525,500	1943-1972	0.9723	0.2983	15	7
	Ohio River at Sewickley	50,480	1943-1972	0.9547	0.2581	8	2
	Tygart Valley River At Colfax	3,529	1940-1969	0.9291	0.2263	7	2
	Tygart Valley River Near Dailey	479	1940-1969	0.8985	0.3324	1	0
Rhine	Rhine at Lobith	160,800	1911-1940	0.9897	0.4254	46	8
	Rhine at Rheinfelden	34,550	1931-1960	0.9715	0	43	9
	Rhine at Domat/Erms	3,229	1911-1940	0.9048	0.4176	7	1
	Emme River at Emmenmatt	443	1915-1944	0.8739	0.3447	7	1

Note: The estimates of long memory parameter  $d$  are adopted from Wang et al. [2007].

- (1) The predictability is closely related to autocorrelations at lag 1, ACF(1). An exponential relationship between ACF(1) and the predictability at different CE levels can be seen by visual inspection at Figure 1.
- (2) The estimates of  $d$  with the *S-MLE* method versus the predictability at two level (SACE = 0.1 and 0.5) are plotted in Figure2 (Note that the zero estimates of  $d$  are removed in the plot). We can discern a exponential relationship between the two, but the relationship is less clearer than that between the predictability and ACF(1).
- (3) There is a positive relation between the predictability and the watershed scale. It is shown in Figure 3a, and Figure 3b that, the larger the watershed scale, the better the predictability. Because it has been found that the relationship between the long-memory parameter  $d$  and the basin scale is very weak [Wang et al., 2007], whereas ACF(1) has a good relationship with the basin scale (shown in Figure 4), therefore, it seems that the increase of predictability with the increase of basin scale mainly results

from the increase of autocorrelation with the increase of basin scale.

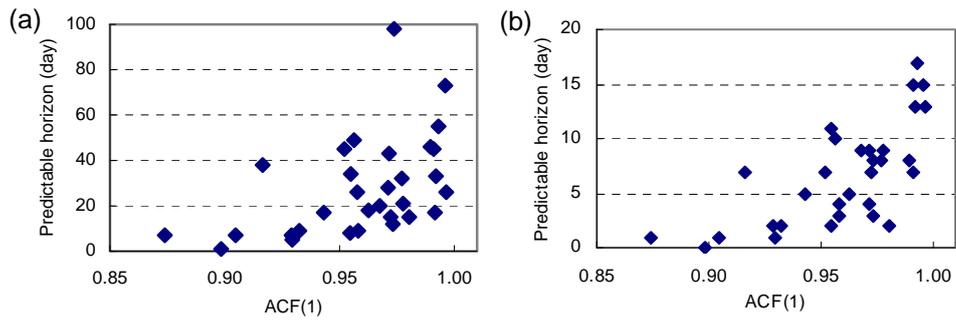


Figure1 Predictability versus ACF(1) for streamflow processes (a) at SACE=0.1 level and (b) SACE = 0.5 level

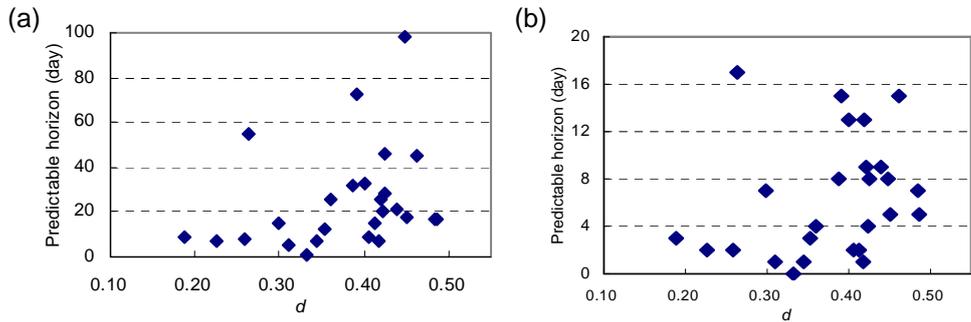


Figure 2 Predictability versus long memory parameter for streamflow processes (a) at SACE=0.1 level and (b) SACE = 0.5 level

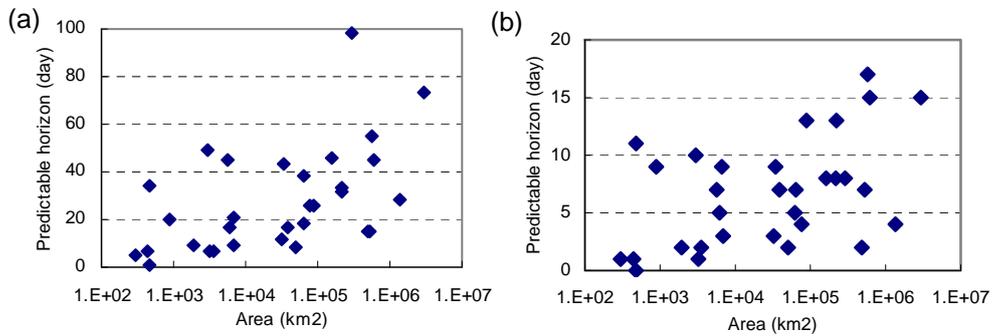


Figure 3 Predictability versus watershed scale for streamflow processes (a) at SACE = 0.1 level and (b) SACE = 0.5 level

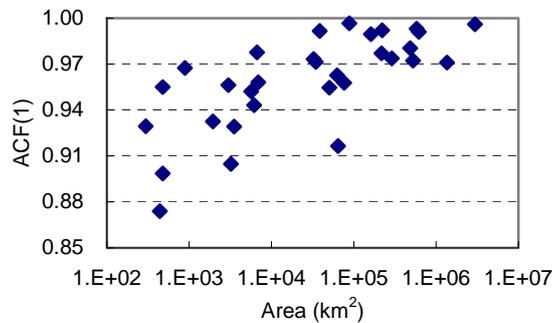


Figure 4 ACF(1) versus watershed scale for streamflow processes

#### 4. UNCERTAINTIES IN THE ESTIMATION OF PREDICTABILITY

Predictability is impacted by many factors, including meteorological factors (such as how river flow is fed, temporal and spatial variability of precipitation processes), and basin characteristics (such as the size, topography, control structures, and drainage network of the basin, and land cover types). These factors may vary from event to event, from season to season, and from region to region, which make the predictability more or less dynamic, rather than an invariable value, albeit it is generally a stable physical feature of a streamflow process.

Apart from the temporal and spatial variability in various factors that impact the predictability, there are several sources of uncertainty in measuring SP of real-world processes, including:

(a) Uncertainty in model selection, especially when different mechanisms may act simultaneously underlying a time series. We suggest using AR model to estimate predictability for the purpose of a “fair” comparison, because of its easiness for using and simplicity for building.

(b) Uncertainty in model parameter estimation, especially when the data size is not big enough. Generally, the larger the data size, the better the parameter estimates because the availability of larger samples allows one to better inspect the asymptotical properties of model parameter.

(c) Uncertainty in data selection. Due to possible long-term variation in climate system and the change of basin characteristics mainly due to human activities, watershed system may exhibit long-term variation. In turn, the mean values of streamflow processes may change over time, thus resulting in the exaggeration of predictability if the mean value of the calibration dataset differs significantly from that of the validation dataset. Therefore, different data selection may give different predictability estimates.

(d) Uncertainty in data quality, which may results in slight exaggeration of predictability. For example, for some gauging stations, the gauged discharges are often the same for two weeks or even over a month continuously, especially when the discharge is very low. For instance, Minnesota River at Ortonville, MN, the discharge kept to be 58 cubic feet per second (cfs) for 17 days (1945.1.12 ~ 1945.1.28); 0.7 cfs for (December 28, 1964 ~ February 5, 1965); 3 cfs for 49 days (1968.1.6 ~ 1968.2.24), 11 cfs for 44 days (1971.1.11 ~ 1971.2.16). While in some cases, this may be true, but most probably this is due to the limited measurement accuracy or even error, which may lead to an slight exaggeration of predictability.

Above uncertainties make it impossible to make precise estimates of predictability, even for the specific approximately stationary period of time that we chose for each streamflow process in the present study.

Due to the seasonality in precipitation and vegetation coverage, streamflow processes usually exhibit strong seasonality, not only in the mean values and variances but also in the autocorrelation structures [see e.g., Wang et al., 2006]. Therefore, the presence of seasonality makes the estimation of predictability of streamflow processes more complicated, and would more or less result in seasonal variation of predictability of streamflow processes. But the seasonality of predictability is not treated in the present study.

#### 5. CONCLUSIONS

In the present study, the predictability of 31 daily average discharge series observed in 8 river basins in Europe and northern America are estimated with univariate time series approach that is based on time series analysis techniques. The results show that, Although the existence of long memory is detected in the daily streamflow processes [Wang et al., 2007], the predictability of these streamflow processes is more similar to that of short-memory AR processes than to that of long-memory ARFIMA processes, indicating that the predictability of the streamflow process is more dominated by short-range autocorrelations than by the existence of long-memory. Possible explanation for that may be in that, the presence of more profound short range memory and seasonality dominate the estimation of predictability albeit the existence of long-memory, which make the effect of long-memory cannot show itself in the predictability measurement. It is also shown that the predictability

is positively related to watershed scale, that is, the larger the watershed scale, the better the predictability of the streamflow process, and this kind of relationship is mainly resulted from positive relationship between autocorrelation and basin scale.

Predictability is impacted by many factors, which make the predictability more or less dynamic, rather than a invariable value. In the present study, the data set were chosen elaborately so as to make the data series approximately stationary and avoid significant impacts of climate change and human interventions (e.g., reservoir impoundment). But many uncertainties still exist, which make it impossible to make precise estimates of predictability.

## ACKNOWLEDGEMENTS

The financial supports from the National Science Foundation of China (Project No. 40771039) and the Research Fund for the Doctoral Program of Higher Education (Project No.20060294003) are gratefully acknowledged.

## REFERENCES

- Andersson, M. K. Do long-memory models have long memory? *International Journal of Forecasting* 16, 121–124, 2000
- Berg, A.A., and Mulroy, K.A., Streamflow predictability in the Saskatchewan/Nelson River basin given macroscale estimates of the initial soil moisture status. *Hydrological Sciences Journal*, 51(4), 642-654, 2006.
- Bloschl, G., and Zehe E., On hydrological predictability. *Hydrol. Process.* 19, 3923–3929, 2005.
- Clements, M. P., and Hendry, D. F., *Forecasting economic time series*. Cambridge University Press, 1998.
- Granger, C.W.J., and Newbold, P., *Forecasting Economic Time Series (2nd)*. Academic Press, Orlando, FL., 1986.
- Hastenrath, S. Predictability of Anomalous River Discharge in Guyana. *Nature*, 345: 53-54, 1990.
- Hosking, J. R. M., Modeling persistence in hydrological time series using fractional differencing, *Water Resour. Res.*, 20(12), 1898-1908, 1984.
- Mandelbrot, B. B., and Wallis J.R., Some long-run properties of geophysical records, *Water Resour. Res.*, 5, 321–340, 1969.
- Maurer, E.P., and Lettenmaier, D. P. Predictability of seasonal runoff in the Mississippi River basin. *J. Geophys. Res.*, 108(D16), 8607, doi: 10.1029/2002JD002555, 2003.
- Mudelsee M., Long memory of rivers from spatial aggregation. *Water Resour. Res.*, 43, W01202, doi:10.1029/2006WR005721, 2007.
- Nash, J. E., and Sutcliffe, J.V., River flow forecasting through conceptual models, I, A discussion of principles, *J. Hydrol.*, 10, 282–290, 1970.
- Raunig, B., The longer-horizon predictability of German stock market volatility. *International Journal of Forecasting*, 22(2), 363-372, 2006.
- Wang, W., van Gelder, P.H.A.J.M., Vrijling, J.K., and Ma, J., Predictability of streamflow processes of the Yellow River. In: *Proceedings of the 6th International Conference on Hydroinformatics*. Singapore : World Scientific, pp.1261-1268, 2004.
- Wang, W., van Gelder P.H.A.J.M., Vrijling J.K., and Ma J., Forecasting Daily Streamflow Using Hybrid ANN Models. *Journal of Hydrology*, 324, 383-399, 2006.
- Wang, W., van Gelder P.H.A.J.M., Vrijling J.K., and Chen X., Detecting long-memory: Monte Carlo simulations and application to daily streamflow processes. *Hydrol. Earth Syst. Sci.*, 11, 851-862, 2007.