

**УДК 556.16.45 + 523.745**

**Артёмченко В.А, Петрович В.В., канд. техн. наук**

**МЕТОД ВОССТАНОВЛЕНИЯ СВЕРХБОЛЬШИХ ПРОПУСКОВ В  
ГИДРОЛОГИЧЕСКИХ ВРЕМЕННЫХ РЯДАХ**

**Анотація.** Наведений метод відновлення безперервних пропусків в гідрологічних рядах даних спостережень.

Об'єкт дослідження: ряди середньодобових витрат води 4 річок одного басейну.

Мета роботи: усунення надвеликих пропусків в рядах гідрологічних даних.

Помилка відновлення ряду даним методом складає менш ніж 15%.

**Ключові слова:** гідрологічні часові ряди, відновлення безперервних пропусків в рядах даних, помилка відновлення гідрологічного ряду.

**Аннотация.** Приведен метод восстановления непрерывных пропусков в гидрологических рядах данных наблюдений.

Объект исследования: ряды среднесуточных расходов воды 4 рек одного бассейна.

Цель работы: устранение сверхбольших пропусков в рядах гидрологических данных.

Ошибка восстановления ряда данным методом составила менее 15%.

**Ключевые слова:** гидрологические временные ряды, восстановление непрерывных пропусков в рядах, ошибка восстановления гидрологического ряда.

**Annotation.** In Article is brought method of the reconstruction unceasing skipping in hydrological time series.

Object of studies are the series of the daily consumption of water of one catchment basin.

The aim of the research is a eliminating, very – large – scale skipping in hydrological data.

Error of the reconstruction with given method forms less than 15%.

**Keywords:** hydrological time series, reconstruction unceasing skipping in hydrological time series, error of the time series reconstruction.

## Введение

Для получения адекватного прогноза исходные гидрологические ряды должны быть высокого качества (без пропусков и ошибочных значений).

В противном случае процедура может не выявить закономерностей в поведении ряда, которые в дальнейшем используются непосредственно для целей прогнозирования.

Если говорить только о небольших пропусках в исходных гидрологических рядах ( несколько идущих подряд значений), то такие пропуски могут быть легко устранены при помощи обычных методов интерполирования, причем с достаточно малыми ошибками.

Однако в данной работе будем рассматривать очень большие непрерывные пропуски величиной порядка 365 значений, или даже, например, 5х365 значений.

Устранение таких сверхбольших пропусков в гидрологических рядах представляет собой нетривиальную задачу, поскольку ее решение не может быть сведено к применению обычных процедур интерполяции.

Как известно, при недостаточной гидрологической характеристике объекта прогнозирования часто используются данные по рекам-аналогам, что требует совместного анализа поведения временных рядов соответствующих рек [1,2].

Ключевая идея данного метода восстановления сверхбольших пропусков также основывается на том, что реки, принадлежащие одному и тому же бассейну, должны иметь взаимосвязанное поведение в своих расходах.

Установив конкретные взаимосвязи в расходах, можно далее восстановить и сверхбольшие пропуски в рядах исходных данных.

В исследовании использовалась векторно-матричная нотация, принятая в системе **MATLAB** и ее аналогах.

### Модель процесса

Пусть, например, есть 4 реки (реки  $R_1 \dots R_4$ ), принадлежащие одному бассейну ( в данном случае это реки бассейна р. Десна).

При этом учитывалось, что линейная связь между стоком двух сравниваемых рек, оцениваемая коэффициентом корреляции  $r \geq 0.7$ , является обязательным условием при выборе реки-аналога (СНиП 2.01.14-83).

Для этих рек имеются непрерывные (без пропусков) данные наблюдений за среднесуточными расходами,  $\text{м}^3/\text{с}$ , в течение нескольких десятилетий, начиная с 1.01.1960 г. и по 31.12.2009 г.

Для простоты вычислений данные за 29 февраля высокосных годов не рассматривались (считалось, что в году 365 дней).

Теперь положим, что у одной из этих рек (например, реки  $R_4$ ) имеется большой пропуск в рядах данных, в то время как для трёх других рек имеем данные без пропусков.

Задача состоит в восстановлении данных за последние 5 лет наблюдений (от 16426 и до 18250 значений ряда) для реки  $R_4$ .

На первом этапе исследований по данным за 1960-2004 г.г. (данные для всех рек  $R_1 \dots R_4$  без пропусков) устанавливали функциональную зависимость вида

$$R_4 = \Psi (R_1; R_2; R_3), \quad (1)$$

где  $\Psi$  – векторная функция (векторных) аргументов,

$R_1, R_2, R_3, R_4$  – вектора (ряды) среднесуточных расходов соответствующих рек,  $\text{м}^3/\text{с}$ ,

что позволило построить 7 моделей процесса.

При этом считали, что взаимосвязи между расходами всех рек линейные (см. выше).

Для всех моделей определяли коэффициенты, которые представляли собой обычные скалярные величины (вычислялись во всех случаях по методу наименьших квадратов).

**Таблица 1** - Ниже в таблице приведены виды моделей и полученные коэффициенты.

Номер модели	Вид функциональной зависимости
1.	$R4 = K1 \cdot R1$ , $K1 = 8.8103$
2.	$R4 = K2 \cdot R2$ , $K2 = 6.5587$
3.	$R4 = K3 \cdot R3$ , $K3 = 2.0897$
4.	$R4 = K1 \cdot R1 + K2 \cdot R2$ , $K1 = -0.7119$ ; $K2 = 6.9578$
5.	$R4 = K1 \cdot R1 + K3 \cdot R3$ , $K1 = 0.7605$ ; $K3 = 1.9700$
6.	$R4 = K2 \cdot R2 + K3 \cdot R3$ , $K2 = 2.4672$ ; $K3 = 1.4227$
7.	$R4 = K1 \cdot R1 + K2 \cdot R2 + K3 \cdot R3$ , $K1 = -1.6583$ ; $K2 = 3.2729$ ; $K3 = 1.4658$

Как видно, модели с номерами 1, 2 и 3 рассматривают функцию одного аргумента, модели с номерами 4, 5 и 6 – функцию двух аргументов, модель с номером 7 – функцию трех аргументов.

Во всех случаях нахождение коэффициентов полностью определяло вид функции (1).

Далее, пользуясь этими моделями, восстанавливали среднесуточные расходы для реки R4 за последние 5 лет наблюдений (1825 значений) и строили графики для реальных и восстановленных значений данных за 2005-2009 годы.

Как оказалось, наилучшие результаты восстановления дают модели 3 и 5 (но это не всегда однозначно отражается на величине ошибки восстановления участка ряда).

Итоговый результат восстановления значений ряда может представлять собой среднее арифметическое, полученное в результате расчетов по этим двум моделям.

На рис.1 и рис.2 приведены соответствующие графики расходов, построенные по данным моделей 3 и 5.

### **Измерение ошибки восстановления пропущенных значений**

Как было отмечено выше, при численном эксперименте из исходных данных реки R4 искусственно удаляли значения ряда за последние 5 лет наблюдений.

Затем выполняли восстановление пропусков и сравнивали полученные результаты с реальной действительностью.

Такой подход дает возможность достаточно точно оценить ошибку восстановления пропущенных значений.

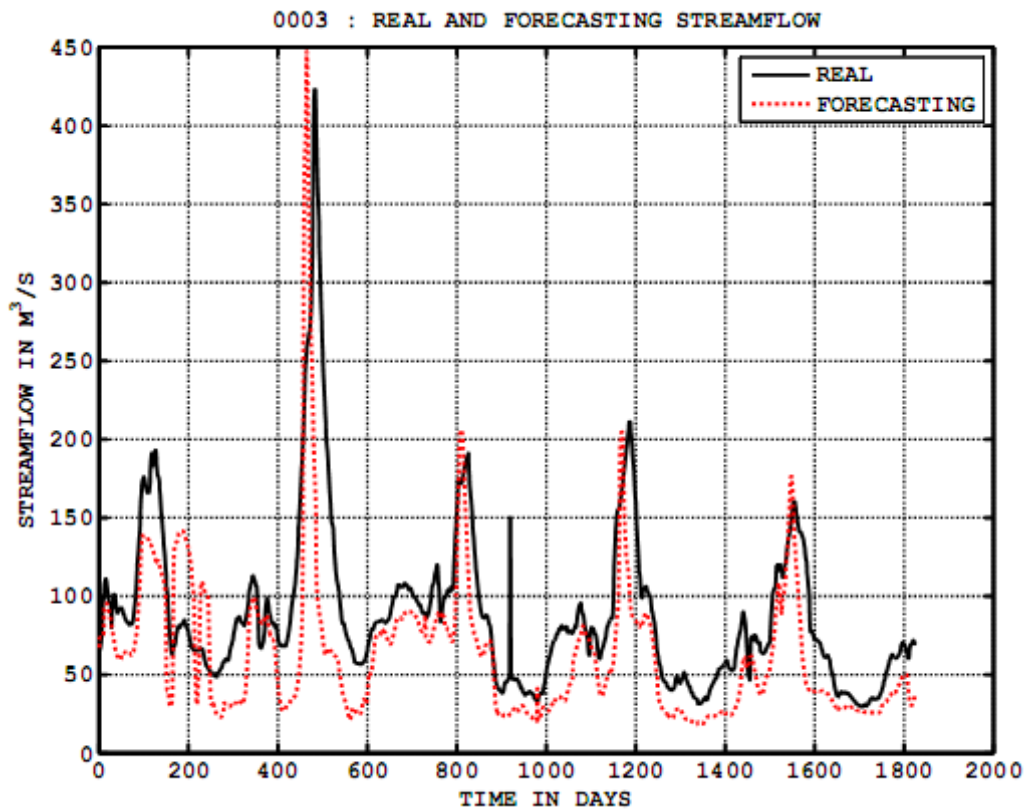


Рисунок 1

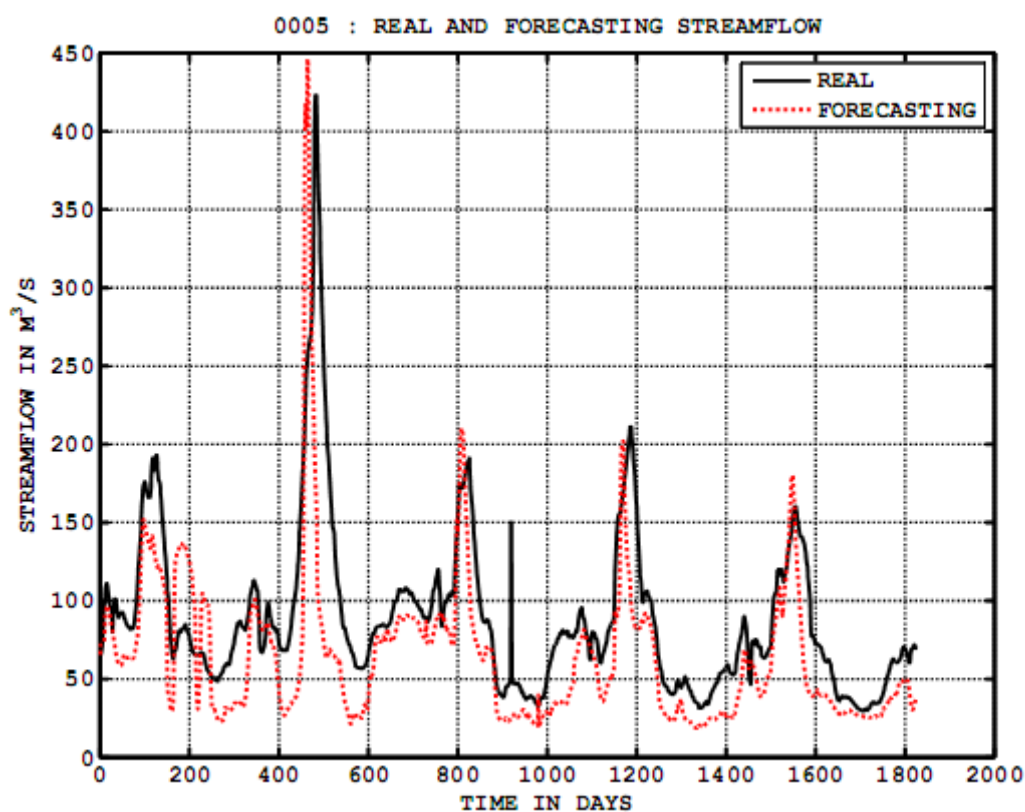
В нотации **MATLAB** ошибка **E** восстановления ряда определяется в процентах как

$$E = 100. * \text{SUM}(\text{ABS}(\text{R}-\text{F}))./(\text{L} * \text{RANGE}(\text{R})) \quad (2)$$

и вычисляется при одновременном выполнении двух условий :

$$\begin{cases} \text{L} = \text{LENGTH}(\text{R}) = \text{LENGTH}(\text{F}) > 1, \\ \text{RANGE}(\text{R}) \neq 0. \end{cases}$$

Понятно, что множитель 100 появляется в (2), поскольку ошибку выражаем в процентах.



**Рисунок 2**

Здесь  $E$  – скаляр;

$R$  – вектор (ряд) истинных величин;

$F$  – вектор (ряд) восстановленных (спрогнозированных) величин;

$L$  – длина вектора  $R$  (вектора  $F$ ), скалярная величина;

$(R-F)$  – разница между реальными и восстановленными значениями. В данном случае

$(R-F)$  – вектор (ряд), отдельные составляющие которого могут быть как положительными, так и отрицательными числами

$RANGE(R) = MAX(R) - MIN(R)$  – размах ряда  $R$ , величина скалярная;

$ABS(R-F)$  – вектор (ряд), отдельные составляющие которого не могут быть отрицательными числами;

$SUM(ABS(R-F))$  – скаляр, представляющий собой сумму всех отклонений (не отрицательных) между  $R$  и  $F$  по всей длине этих рядов ( $R$  и  $F$ ).

По сути, ошибка восстановления уже определяется выражением  $SUM(ABS(R-F))$ , но для существенно различных рядов в общем случае могут получаться несравнимые результаты.

Разделим далее последнее выражение на длину ряда  $R$  (или длину ряда  $F$ ).

Теперь можно сравнивать ошибки, полученные при восстановлении рядов различной длины.

Затем делим все на разницу между максимальным и минимальным значением ряда  $R$ .

В итоге появляется возможность адекватного сравнения ошибок восстановления для рядов с совершенно различным размахом (см. (2)).

Таким образом определяли ошибки восстановления для участка ряда за последние 5 лет.

Ошибка восстановления  $E$  для всех случаев составила менее 15% (11,9111% для модели 1; 12,3582% для модели 2; 8,1656% - модели 3; 12,5898% - модели 4; 7,9036% - модели 5; 8,3533% - модели 6 и 8,9513% - модели 7).

### **Выводы**

1. Восстановление сверхбольших непрерывных пропусков в гидрологических рядах данным методом вполне реально.

Во всех случаях пиковые значения расходов и ход кривых на графиках описываются с достаточно большой точностью.

При малых значениях расходов точность восстановления снижается. Однако, поскольку практический интерес вызывают именно наибольшие расходы, некоторое снижение точности восстановления в период малых расходов не столь существенно.

2. Ошибка восстановления для всех рассмотренных случаев (модели 1...7) составила менее 15%.

3. Данные линейные модели целесообразно выбирать за некий эталон, по которому в дальнейшем можно оценивать качество работы более совершенных нелинейных моделей.

### **Литература**

1. Пособие по определению расчетных гидрологических характеристик. – Л. : Гидрометеиздат. – 1984. – 448 с.
2. Рождественский А.В. Оценка точности гидрологических расчетов / А.В. Рождественский, А.В. Ежов, А.В. Сахарюк // Л. : Гидрометеиздат, – 1990. – 276 с.