

УДК 639.053.1

ВЕРОЯТНОСТНАЯ МОДЕЛЬ
КОМПЛЕКСА БИОГИДРОЛОГИЧЕСКИХ ХАРАКТЕРИСТИК
АЗОВСКОГО БАССЕЙНА

В.Г.Дубинина, А.М.Березовский, И.В.Попов
(АзНИИРХ)

Настоящая работа представляет собой развитие исследования, выполненного ранее одним из авторов (Дубинина, 1969). Цель работы — изучить возможность применения различных регрессионных моделей для прогноза промышленного возврата донского судака на основе знания комплекса гидрометеорологических элементов водоема и выбрать оптимальную модель с точки зрения точности и количества факторов, необходимых для составления прогноза.

Линейные модели. Рассмотрим первый вариант модели в виде линейной комбинации абиотических факторов

$$y = a_0 + \sum_{i=1}^{13} a_i x_i, \quad (I)$$

- где y — промышленный возврат судака, млн. шт. ;
 x_1, x_2, x_3 — сток Дона соответственно за время захода производителей на нерест, за время нереста ($9^0 \leq t \leq 19,5^0$), за время ската личинок и производителей, км³ ;
 x_4 — время затопления поймы с учетом благоприятной для нереста температуры ($9^0 \leq t \leq 19,5^0$);
 x_5 и x_6 — дата устойчивого перехода суточной температуры воздуха через 6^0 и через 10^0 ;
 x_7 — средняя температура воздуха за зиму ;

- X_8 - коэффициент вариации минимальной температуры воздуха за время нереста;
- X_9 - число производителей, идущих на нерест;
- X_{10} - соленость воды восточной части Таганрогского залива летом (июнь, июль, август), ‰;
- X_{11} - число дней с ветром ≥ 8 м/сек за летний период над экваторией Азовского моря;
- X_{12} - сумма среднесуточных температур воды в Дону от 1° до 9° (за время захода производителей на нерест), град.;
- X_{13} - сумма среднесуточных температур за период осеннего понижения от 9° до 1° .

Обоснование выбора этих факторов в качестве исходных для построения модели приведено в работе В.Г.Дубининой (1969).

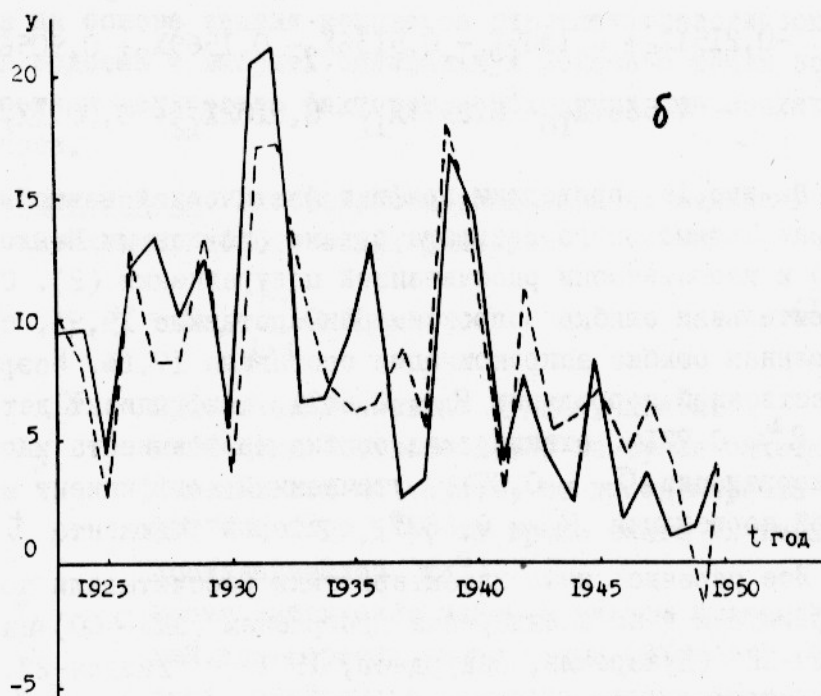
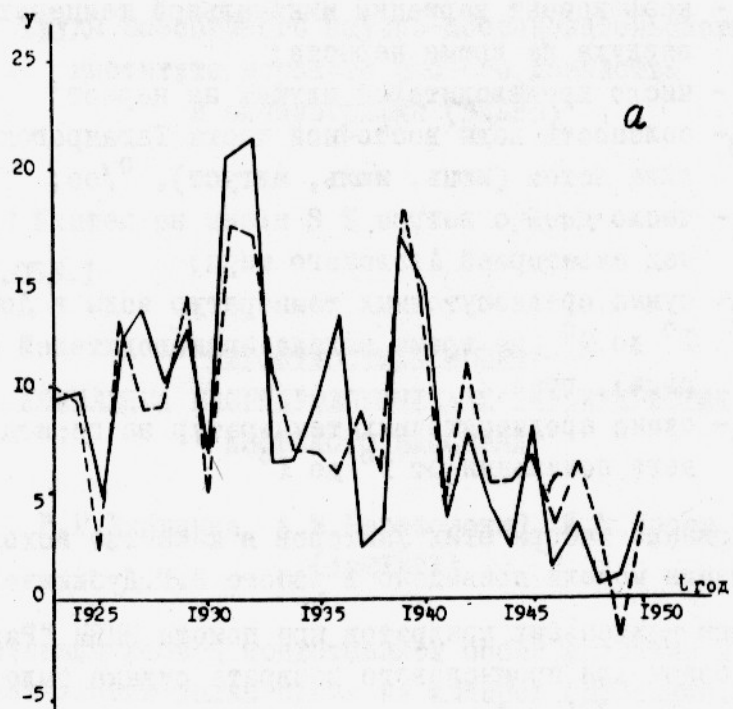
Методом наименьших квадратов при помощи ЭЦВМ "Раздан-2" уравнение связи для промыслового возврата судака было получено в виде:

$$\begin{aligned}
 Y = & 22,48I_3 - 0,6594X_1 - 0,5089X_2 + 0,4085X_3 + 0,1208X_4 - \\
 & - 0,2151X_5 + 0,1365X_6 - 0,5136X_7 - 0,1569X_8 + 0,9056X_9 - (2) \\
 & - 7,5664X_{10} - 0,2973X_{11} - 0,0102X_{12} + 0,0071X_{13}.
 \end{aligned}$$

На рис.1а, приведены графики фактической зависимости от времени промыслового возврата судака (по данным Бойко Е.Г. 1955) и теоретически рассчитанной по уравнению (2). Средняя относительная ошибка аппроксимации составила 19,9%, средневзвешенная ошибка аппроксимации составила 17,1%. Коэффициент множественной корреляции $R = 0,8494$, коэффициент детерминации $R^2 = 0,7214$, стандартная ошибка коэффициента множественной корреляции $\sigma_R = 0,0744$, уточненный коэффициент множественной корреляции $R = 0,6803$, критерий Стьюдента $t = 11,41$.

Все перечисленные характеристики рассчитывали по известным формулам и по стандартным программам (АКИ-400) на ЭЦВМ "Минск-22" (Дукарский, Закурдаев, 1971) и "Раздан-2". Так, критерий Стьюдента вычислялся в виде

$$t = \frac{R}{\sigma_R}$$



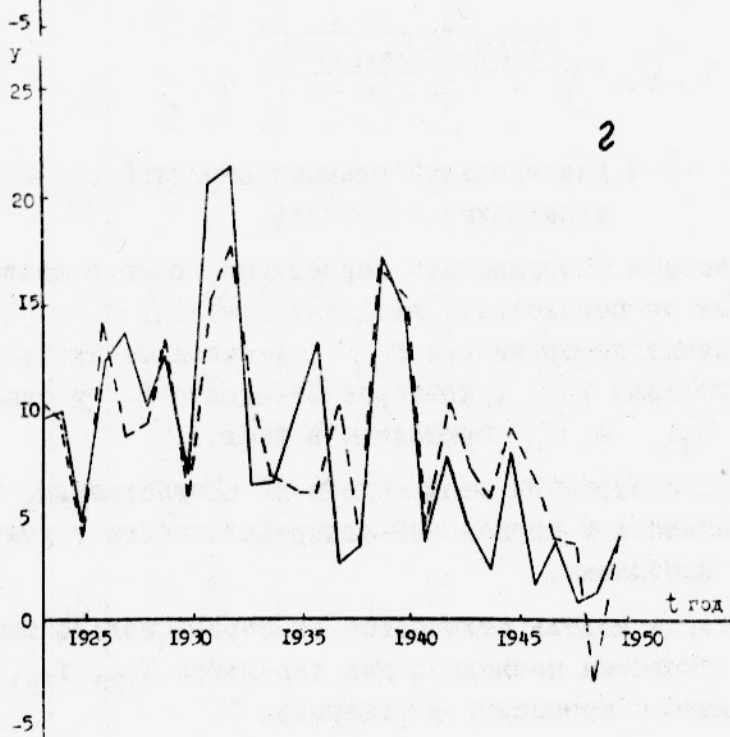
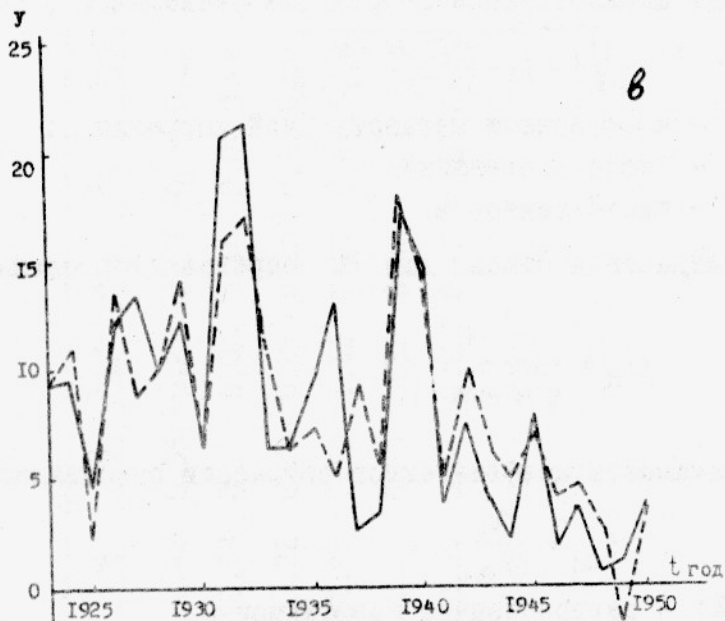


Рис. I. Теоретическая (---) и фактическая (—) зависимости промышленного возврата судака от времени:
 а - уравнение (2); б - уравнение (3);
 в - уравнение (4); г - уравнение (5).

где R и σ_R имеют оговоренное выше значение. Уточненный коэффициент множественной корреляции вычислялся в виде

$$\bar{R} = \sqrt{1 - (1 - R^2) \frac{N-1}{N-m-1}},$$

где R - коэффициент множественной корреляции;

N - число наблюдений;

m - число факторов.

Стандартная ошибка для R определялась согласно выражению

$$\sigma_R = \frac{1 - R^2}{\sqrt{N - m - 1}}.$$

Значимость коэффициентов регрессии оценивалась в виде

$$t_{a_i} = \frac{a_i}{\sigma_{a_i}},$$

здесь a_i - регрессионные коэффициенты;

σ_{a_i} - стандартные ошибки a_i .

$$\sigma_{a_i} = \sqrt{\frac{\sum (Y_{ф.} - Y_{расч.})^2 C_{ii}}{N - m - 1}},$$

где C_{ii} - i -й диагональный элемент обратной матрицы системы нормальных уравнений.

Частные коэффициенты корреляции прямого возврата судака с каждым из параметров, входящих в модель γ_{y_i} , ($i = 1, 2, \dots, 13$), коэффициент детерминации $\gamma_{y_i}^2$, уточненные частные коэффициенты корреляции $\bar{\gamma}_{y_i}$, критерии Стьюдента t и стандартные ошибки $\sigma_{\gamma_{y_i}}$ и σ_{a_i} приведены в табл. I.

При построении этой модели мы предполагали, что условия независимости и нормальной распределенности учитываемых параметров выполнены.

Вторая модель отличается от первой только тем, что в нее не включены последние два параметра X_{12} , X_{13} , слабо коррелирующие с промышленным возвратом Y .

Таблица I

Частные коэффициенты корреляции и стандартные ошибки для модели,
описываемой уравнением (2)

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}
r_{xy}	-0,18	-0,37	0,12	0,20	-0,40	0,34	-0,26	-0,42	0,35	-0,28	-0,37	-0,07	0,15
r^2_{xy}	0,04	0,14	0,01	0,04	0,16	0,11	0,07	0,18	0,12	0,08	0,14	0,01	0,02
$\sigma_{r_{xy}}$	0,26	0,23	0,26	0,26	0,22	0,24	0,25	0,22	0,23	0,25	0,23	0,26	0,02
t	0,73	1,61	0,46	0,78	1,80	1,42	1,07	1,93	1,50	1,15	1,65	0,26	0,58
\bar{r}_{xy}	0	0,27	0	0	0,32	0,22	0,07	0,34	0,25	0,12	0,28	0	0
σ_{a_i}	0,92	0,34	0,90	0,16	0,13	0,10	0,49	0,09	0,64	0,68	0,19	0,03	0,001

Уравнение, найденное методом наименьших квадратов, в этом случае имеет вид

$$Y = 27,17102 - 0,92003X_1 - 0,60004X_2 + 0,60130X_3 + \\ + 0,13371X_4 - 0,21185X_5 + 0,12904X_6 - 0,54168X_7 - \\ - 0,16549X_8 + 0,92755X_9 - 8,45036X_{10} - 0,30767X_{11}.$$

Соответствующая кривая зависимости промыслового возврата судака от времени представлена на рис.1б. Средняя относительная и средневзвешенная погрешности аппроксимации составляют соответственно 26 и 13,6%. Коэффициент множественной корреляции $R = 0,8438$, $R^2 = 0,7120$, $\sigma_R = 0,0720$, $\bar{R} = 0,7170$, $t = 11,72$. Частные коэффициенты корреляции и относительные ошибки представлены в табл.2. Как и в предыдущем случае, условия применимости регрессионного анализа специально не проверялись, но предполагались выполненными.

Таблица 2

Частные коэффициенты корреляции и стандартные ошибки для модели, описываемой уравнением (3)

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
r_{xy}	0,33	0,46	0,19	0,22	0,41	0,32	0,31	0,44	0,45	0,32	0,38
r^2_{xy}	0,11	0,22	0,04	0,05	0,17	0,10	0,10	0,20	0,20	0,10	0,15
$\sigma_{r_{xy}}$	0,22	0,19	0,24	0,24	0,21	0,22	0,22	0,20	0,20	0,22	0,21
t	1,50	2,40	0,79	0,93	1,96	1,42	1,40	2,21	2,23	1,41	1,80
\bar{r}_{xy}	0,23	0,40	0	0	0,34	0,21	0,21	0,38	0,38	0,21	0,31
σ_{a_i}	0,65	0,28	0,78	0,15	0,12	0,10	0,41	0,08	0,46	0,28	0,18

В третью линейную регрессионную модель

$$y = a_0 + \sum_{i=1}^{11} a_i x_i$$

были включены следующие параметры:

- X_1 - суммарный сток Дона за время хода производителей на нерест, время нереста и время ската;
- X_2 - время затопления поймы с учетом благоприятной для нереста температуры ($9^0 \leq t \leq 19,5^0$);

- X_3 - дата устойчивого перехода минимальной суточной температуры воздуха через 6° ;
- X_4 - дата устойчивого перехода среднесуточной температуры воздуха через 10° ;
- X_5 - средняя температура воздуха за зиму;
- X_6 - коэффициент вариации минимальной температуры воздуха за время нереста;
- X_7 - число производителей, идущих на нерест;
- X_8 - соленость воды восточной части Таганрогского залива летом;
- X_9 - число дней с ветром ≥ 8 м/сек за летний период над экваторией Азовского моря;
- X_{10} - сумма среднесуточных температур воды в Дону от 1° до 9° (за время захода производителей на нерест);
- X_{11} - сумма среднесуточных температур воды за период осеннего понижения от 9° до 1° .

Уравнение связи промыслового возврата судака с абиотическими факторами перечисленного типа имеет вид:

$$\begin{aligned}
 Y = & 20,4316 - 0,359X_1 + 0,0816X_2 - 0,2498X_3 + 0,1511X_4 - \\
 & - 0,4755X_5 - 0,1252X_6 + 0,8269X_7 - 7,8932X_8 - 0,2969X_9 - \\
 & - 0,0137X_{10} + 0,0127X_{11}. \quad (4)
 \end{aligned}$$

Соответствующий этому уравнению график представлен на рис. 1в. Средняя относительная ошибка аппроксимации составила $\xi_1 = 1\%$, средневзвешенная $\xi_2 = 15,3\%$. Коэффициент множественной корреляции $R = 0,83$, детерминации $R^2 = 0,7$, $\sigma_R = 0,07$, $t = 11,29$, $R = 0,7$. Частные коэффициенты корреляции и относительные ошибки представлены в табл. 2. Условия применимости регрессионного анализа предполагались выполненными.

Четвертая, и последняя, линейная модель отличается от предыдущей только тем, что в нее не включены параметры X_{10} и X_{11} . Уравнение связи при этом приобретает следующий вид:

$$Y = 28,5 - 0,4656X_1 + 0,0993X_2 - 0,2727X_3 + 0,1539X_4 - 0,4758X_5 - 0,1241X_6 + 0,8466X_7 - 9,3397X_8 - 0,3034X_9. \quad (5)$$

На рис. 1г представлено графическое изображение зависимости $y(t)$, соответствующее этому уравнению. Средняя относительная ошибка аппроксимации в этом случае $\xi_1 = 6,5\%$, средневзвешенная $\xi_2 = 17,3\%$, $R = 0,82$, $R^2 = 0,66$, $\sigma_R = 0,08$, $t = 10,29$, $\bar{R} = 0,7$. Все остальные характеристики включены в табл. 4.

Таблица 3

Частные коэффициенты корреляции и стандартные ошибки для модели, описываемой уравнением (4)

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
r_{xy}	0,30	0,15	0,46	0,37	0,25	0,37	0,33	0,31	0,38	0,13	0,30
r^2_{xy}	0,09	0,02	0,21	0,14	0,06	0,14	0,10	0,10	0,14	0,01	0,09
$\sigma_{r_{xy}}$	0,23	0,24	0,19	0,21	0,23	0,21	0,22	0,22	0,21	0,25	0,23
t	1,30	0,60	2,30	1,70	1,06	1,70	1,46	1,39	1,77	0,51	1,32
\bar{r}_{xy}	0,18	0	0,40	0,29	0,06	0,28	0,22	0,20	0,30	0	0,18
σ_{a_i}	0,29	0,13	0,12	0,09	0,46	0,07	0,59	0,59	0,18	0,03	0,001

Таблица 4

Частные коэффициенты корреляции и стандартные ошибки для модели, описываемой уравнением (5)

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
r_{xy}	0,37	0,18	0,49	0,37	0,35	0,26	0,41	0,35	0,37
r^2_{xy}	0,14	0,03	0,25	0,13	0,12	0,07	0,17	0,12	0,14
$\sigma_{r_{xy}}$	0,20	0,23	0,17	0,20	0,20	0,22	0,19	0,21	0,20
t	1,80	0,79	2,80	1,80	1,70	1,20	2,11	1,69	1,80
\bar{r}_{xy}	0,29	0	0,45	0,29	0,26	0,13	0,36	0,27	0,30
σ_{a_i}	0,27	0,13	0,11	0,09	0,40	0,07	0,44	5,87	0,18

Нелинейные модели. Естественным развитием линейных моделей, очевидно, является переход к нелинейным вариантам связи промышленного возврата донского судака с гидрометеорологическими характеристиками бассейна. Исходными параметрами для синтеза таких моделей были приняты следующие:

- Y - промышленный возврат судака в $i + 3$ -й год, млн.шт.;
- X_1 - объем речного стока в период нереста i -го года при $9^0 \leq t \leq 19,5^0$, км³;
- X_2 - объем весеннего стока Дона в i -й год, км³;
- X_3 - расход воды в i -й год нереста при $9^0 \leq t \leq 19,5^0$, м³/сек;
- X_4 - время затопления нерестилищ в i -й год, сутки;
- X_5, X_6, X_7 - среднегодовая площадь ареала в i -й год соответственно 0 ± 9 , 0 ± 10 и $0 \pm 11\%$, км²;
- X_8 - средняя температура воздуха за период нереста;
- X_9 - среднее значение суточных размахов температуры воздуха за время нереста;
- X_{10} - количество дней с благоприятной для нереста температурой воздуха.

Использовались данные за период времени с 1923 по 1942г.

Чтобы проверить условия применимости регрессионного анализа, предварительно для всех параметров были просчитаны основные числовые статистические характеристики: m - математическое ожидание, σ - среднеквадратичное отклонение, s^2 - выборочная дисперсия (смещенная оценка), \bar{s}^2 - выборочная дисперсия (несмещенная оценка), M_3 - несмещенный момент 3-го порядка, M_4 - несмещенный момент 4-го порядка, A - коэффициент асимметрии (несмещенная оценка), E - эксцесс (несмещенная оценка), C - коэффициент вариации в % (табл.5). Кроме того, с этой же целью были просчитаны парные коэффициенты корреляции (табл.6) с учетом их значимости по критерию t и \bar{z} - преобразования Фишера. Анализ этих данных показал, что далеко не все взятые параметры удовлетворяют условиям нормальности и независимости. Для дальнейшей работы с моделью были оставлены факторы X_3 , X_7 , X_8 , X_9 . На рис.2 представлена диаграмма, поясняющая процедуру выбора факторов для включения в модель (прямыми линиями обозначены связи между параметрами). Очевид-

но, параметры X_7, X_8 могут быть сразу же включены в любую модель на базе выбранных десяти факторов как некоррелирующие ни с одним из них (в том числе и между собой). Остальные параметры выбираются исходя из некоррелированности с другими параметрами, входящими в ту или иную модель. Всего возможных вариантов сочетаний некоррелированных параметров получилось девять:

- 1) X_1, X_6, X_7, X_8, X_9 ; 2,3) X_2, X_6, X_7, X_8, X_9 (или X_{10});
 4,5) X_3, X_7, X_8, X_9 (или X_{10}); 6,7) X_4, X_7, X_8, X_9 (или X_{10});
 8,9) X_5, X_7, X_8, X_9 (или X_{10}).

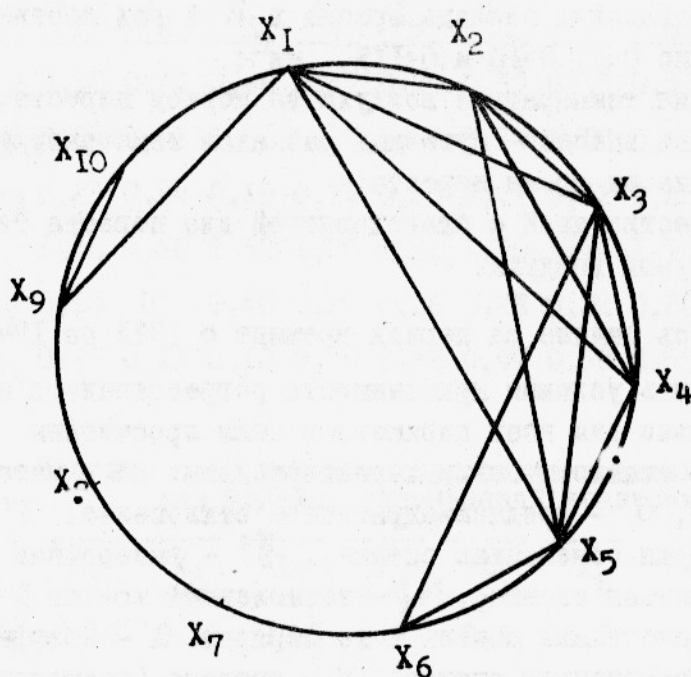


Рис.2. Диаграмма статистических связей параметров нелинейной регрессионной модели.

Для построения модели был выбран четвертый вариант, параметры которого наиболее значимы в процессе формирования урожайности поколений. Методом наименьших квадратов было построено несколько нелинейных моделей $Y = f(X_3, X_7, X_8, X_9)$, из которых наилучшей, в смысле минимума ошибки аппроксимации, оказалась следующая:

Таблица 5

Основные статистические характеристики параметров нелинейной модели
промыслового возврата судака

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
m	12,6	19,6	3441,1	32,9	5844,4	12972,2	27388,8	15,4	11,4	4241
σ	9,0	10,2	2085,8	15,8	1947,7	9954,4	1154,4	11581,5	1,25	1,06
s^2	86,2	111,0	$460 \cdot 10^4$	266,1	$401 \cdot 10^4$	$104 \cdot 10^6$	$142 \cdot 10^6$	1,66	2,57	120,10
\bar{s}^2	81,4	104,9	$435 \cdot 10^4$	251,4	$279 \cdot 10^4$	$99 \cdot 10^6$	$134 \cdot 10^6$	1,57	2,44	113,40
M_3	443,1	658,9	$926 \cdot 10^7$	-3229	$176 \cdot 10^8$	$15 \cdot 10^{11}$	$-169 \cdot 10^{10}$	1,15	-0,786	-7,58
M_4	13357,8	24340,4	$626 \cdot 10^{11}$	$191 \cdot 10^3$	$125 \cdot 10^{12}$	$34 \cdot 10^{15}$	$45 \cdot 10^{15}$	7,18	15,90	$27 \cdot 10^3$
A	0,55	0,56	0,93	-0,74	2,19	1,42	0,99	-0,54	-0,18	-0,005
E	-1,04	-0,8	0,31	0,012	5,96	0,51	-0,51	-0,09	-0,33	-0,89
C	73,7	53,7	57,3	49,5	34,3	78,9	43,6	8,3	14,05	26,02

Таблица 6

Парные коэффициенты корреляции параметров
нелинейной модели

	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀
Y	I	0,42	0	0,45	0	0	0	0	0	-0,41	0
X ₁		I	0,98	0,94	0,75	0,47	0	0	0	-0,42	0
X ₂			I	0,97	0,77	0,53	0	0	0	0	0
X ₃				I	0,61	0,63	0,45	0	0	0	0
X ₄					I	0,84	0,57	0	0	0	0
X ₅						I	0,61	0	0	0	0
X ₆							I	0	0	0	0
X ₇								I	0	0	0
X ₈									I	0	0
X ₉										I	-0,58
X ₁₀											I

$$\begin{aligned}
 Y = & 274,2245 - 32,13987X_8 - 2,68970X_9 - 0,7146547 \cdot 10^{-6} \cdot X_3^2 + \\
 & + 1,63878 \cdot 10^{-8} X_7^2 + 1,132298 X_8^2 + 2,0494412 \cdot 10^{-7} X_3 \cdot X_7 - \\
 & - 1,263487 \cdot 10^{-4} X_7 \cdot X_8 + 0,6482861 \cdot 10^{-4} X_7 \cdot X_9 \quad (6)
 \end{aligned}$$

Соответствующий график зависимости промышленного возврата судака от времени представлен на рис.3. Относительная ошибка аппроксимации составила 26%, средневзвешенная 11% при 95% обеспеченности.

Полученные корреляционные связи и достаточно высокие частные коэффициенты корреляции подтвердили правильность построений и определили группу факторов, наиболее значимых для процесса формирования урожайности поколений судака.

Вычисление частных коэффициентов и проверка рядов по критерию Стьюдента показали малую значимость параметров, характеризующих температуру воды при заходе производителей на

нерест и в период осеннего выхолаживания, которые в последующих расчетах были исключены.

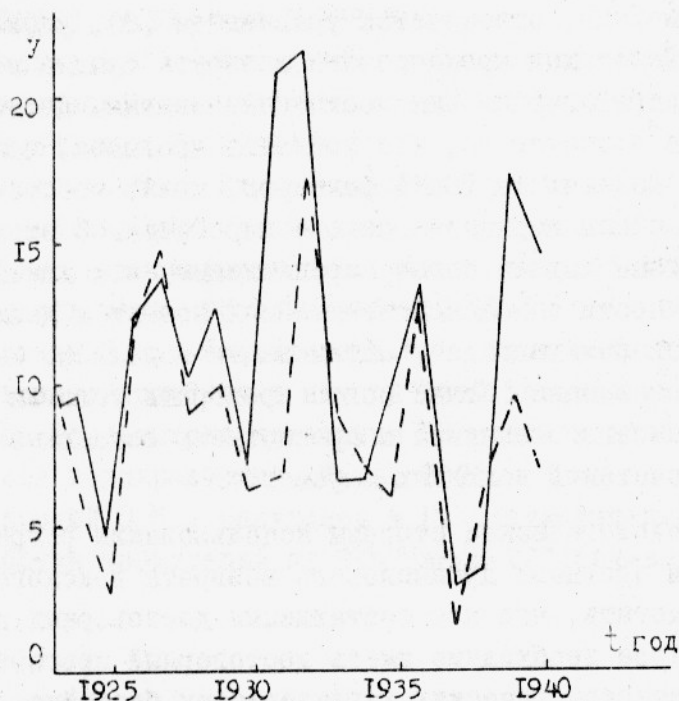


Рис.3. Зависимость промышленного возврата судака от времени:
 - - - - теоретическая кривая, уравнение (6);
 ———— фактическая кривая

Исследованиями установлено, что эффективность размножения рыб на пойме Дона в основном зависит от объема весеннего половодья (r до 0,47), количества идущих на нерест производителей ($r = 0,45$), термического режима (r до 0,44), динамики ($r = 0,38$) и солености вод в восточной половине Таганрогского залива ($r = 0,32$).

Наиболее точной в прогностическом отношении среди линейных моделей оказалась третья, описываемая уравнением (4), $\epsilon_1 = 1\%$, $\epsilon_2 = 15,3\%$. Модель включает II параметров, большая часть из которых представляет собой гидрометеорологические характеристики бассейна в год нереста судака. Возможно, модель будет еще более точной, если тщательно отобразить параметры по критериям взаимной независимости и нормальности.

Неплохие результаты дает и четвертая линейная модель, уравнение (5), $\epsilon_1 = 6,5\%$, $\epsilon_2 = 17,3\%$, включающая всего 9 параметров.

Нелинейная модель, описываемая уравнением (6), удовлетворительно производит ход промыслового возврата судака по годам, но ошибка аппроксимации еще достаточно велика. Достоинством этой модели является то, что точность прогноза, получаемого на линейных моделях из 10-13 факторов, можно обеспечить, используя в нелинейном варианте всего 4 параметра. С этой точки зрения нелинейные модели более перспективны, чем линейные. Для повышения точности описанной нелинейной модели в дальнейшем предполагается решить задачу оптимизации модели по минимуму ошибки аппроксимации, более полно проверить условия применимости регрессионного анализа и просмотреть остальные восемь вариантов сочетаний исходных параметров.

Касаясь методологической стороны использования регрессионных моделей для прогноза промыслового возврата донского судака, следует заметить, что для составления достоверных прогнозов на этой основе необходимо иметь достоверный прогноз комплекса гидрометеорологических характеристик бассейна, входящих в модель. Такой прогноз может быть получен известным методом экстраполяции стационарных случайных процессов. Причем трудности его применения к нестационарным случайным процессам могут быть легко преодолены выделением тренда из исходной реализации с последующим его отдельным анализом функциональными методами.

Полученная связь между промысловым возвратом судака и рядом биотических и абиотических факторов была использована для определения оптимальных для рыбного хозяйства параметров сброса воды через плотину Цимлянского гидроузла (Дубинина, Козлитина, 1971). Описанные модели могут быть с успехом использованы для прогноза промыслового возврата не только донского судака, но и других полупроходных и проходных рыб Азовского бассейна. Очевидно, также, что аналогичные регрессионные модели могут быть построены и для всех остальных промысловых рыб этого водоема.

Л и т е р а т у р а

- Бойко Е.Г. Эффективность естественного размножения и основные пути воспроизводства судака Азовского моря. Труды ВНИРО, т.31, вып.2, 1955.
- Дубинина В.Г. Использование множественной корреляции для прогноза промыслового возврата донского судака. Труды ВНИРО, т.67, вып.1, 1969.
- Дубинина В.Г. Гидрологический режим поймы Нижнего Дона и проблема рыбохозяйственного использования водных ресурсов реки. Автореферат диссертации на соискание уч. степ. канд. геогр. наук. Ростов-на-Дону, 1969.
- Дубинина В.Г., Козлитина С.В. Опыт оптимизации рыбохозяйственных попусков в нижний бьеф водохранилища. Тезисы доклада всесоюзного совещания "Вопросы комплексного использования водохранилищ". Киев, 1971.
- Дударский О.М., Закурдаев А.Г. Статистический анализ и обработка наблюдений на ЭВМ "Минск-22". Изд-во "Статистика", 1971.

A PROBABLE MODEL OF THE COMPLEX OF BIOHYDROLOGIC
CHARACTERISTICS OF THE AZOV SEA BASIN

V.G.Dubinina, A.M.Beresovsky,
I.V.Popov

S u m m a r y

The paper suggests a prognosis of commercial return of fish based on a probable nature of hydrometeorologic parameters of the Azov Sea basin and statistical structure of their interrelations. Some linear and non-linear regression models of commercial returns which include about thirteen biologically most important abiotic factors in the basin are developed with special reference to the pike-perch from the Don River. The main statistical values for each factor are preliminary estimated and conditions for application of various regression models are tested. The best result has been obtained for a non-linear case where only four abiotic parameters are used. Merits and demerits of the models as well as prospects of their development are analysed.