

06, июнь 2016

УДК 574.5: 574.34.045 + 551.521

Комплексный анализ парной корреляционной связи между значениями фотосинтетически активной солнечной радиации и датами максимальной численности фитопланктона в Учинском водохранилище в весенний период

*Юсупова К. О., студентка
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Прикладная математика»*

*Научный руководитель: Корсак М.Н., к.б.н, доцент
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Прикладная математика»
e9@mx.bmstu.ru*

В предыдущих работах авторы показали, что ежегодно в весенний период в пресноводных водоемах наблюдается такое явление, как «цветение» воды, которое вызвано резким увеличением численности микроводорослей. Ранее было выполнено большое количество исследований, которые позволяют утверждать о том, что основным фактором, влияющим на дату наблюдения этого явления, является количество фотосинтетически активной солнечной радиации (ФАР), приходящей на поверхность водоема. Авторами было доказано, что значение суммарной солнечной радиации за пять дней в середине февраля положительно коррелируют с датой наступления пика «цветения» фитопланктона в водоеме ($r \approx 0,8$), что свидетельствует о наличии сильной зависимости между этими параметрами.

После всех предшествующий исследований остались некоторые вопросы, ответы на которые позволили бы окончательно убедиться в том, что фактор солнечной радиации является определяющим в процессе весеннего развития микроводорослей.

Первый из этих вопросов касается улучшения результатов, т. е. возможно ли, что пятидневный интервал – это не единственный временной период, значения солнечной радиации за который, имеют сильную зависимость с датами пика весеннего «цветения» фитопланктона? Другими словами, можно ли найти такой участок времени, в который мы получим значения коэффициента корреляции больше, чем были получены ранее?

И еще остался не оговоренным вопрос: действительно ли при такой сильной зависимости необходимо рассматривать только интенсивность солнечной радиации в

качестве единственного фактора в весенний период, влияющего на варьирование численности микроводорослей в водоеме?

В данной работе путем проведения ряда расчетов была установлена связь между интегральными значениями ФАР за один полный месяц (значение корреляции для февраля было максимальным из всех рассматриваемых временных периодов) и датой наступления пика «цветения» фитопланктона на каждой из двух станций Учинского водохранилища. Необходимо было доказать, что данная связь максимально близко к действительности отражает ежегодное развитие микроводорослей в весенний период в Учинском водохранилище. Затем, необходимо объяснить, почему в данном исследовании не учитываются другие факторы, которые в той или иной степени влияют на рассматриваемый процесс.

Для такого подробного изучения связи между двумя факторами: датой пика «цветения» фитопланктона и значением ФАР за февраль, необходимо было построить комплексный анализ парной корреляционной связи между ними. Это позволит подтвердить наличие этой связи и то, в какой степени эта связь себя проявляет.

Рассмотрим сначала верхнюю часть Учинского водохранилища: ст. Пестово. Для начала найдем значение коэффициента корреляции (r) – показателя линейной связи между двумя факторами: значениями ФАР за февраль и датой наступления максимальной численности фитопланктона. Значения коэффициента корреляции могут варьировать в диапазоне от -1 до 1. Интерес представляет случай, когда коэффициент корреляции очень близок к единице, так как это свидетельствует о наличии сильной линейной связи между рассматриваемыми параметрами.

Таблица 1

Вычисление линейного коэффициента корреляции r

x – значение календарных суток, когда достигается максимум численности фитопланктона за каждый год на ст. Пестово,

y – значение интегральной фотосинтетически активной солнечной радиации за февраль

Год	x , сут.	y , МДж/м ²	Δx	Δy	Δx^2	Δy^2	$\Delta x \times \Delta y$
1998	131	66.17	8.83	6.55	77.97	42.90	57.84
1999	131	64.83	8.83	5.21	77.97	27.14	46.00
2000	114	56.67	-8.17	-2.95	66.75	8.70	24.10
2001	122	62.16	-0.17	2.54	0.03	6.45	-0.43
2002	111	41.71	-11.17	-17.91	124.77	320.77	200.05
2003	124	66.17	1.83	6.55	3.35	42.90	11.99
Σ	733	357.69			350.83	448.87	339.55
n	6	6					
Среднее	122.17	59.62					

На основе данных таблицы 1 вычислим коэффициент корреляции (r) по формуле

$$r_{x/y} = \frac{\sum \Delta x \Delta y}{\sqrt{\sum \Delta x^2 \sum \Delta y^2}} = \frac{339.55}{\sqrt{350.83 \cdot 448.87}} = \frac{339.55}{396.83} = 0.86.$$

Вычислим также среднеквадратические отклонения (σ_x, σ_y), которые могут нам понадобиться далее для вычисления коэффициента регрессии:

$$\sigma_x = \sqrt{\frac{\sum \Delta x^2}{n-1}} = 8.38 \text{ сут.},$$

$$\sigma_y = \sqrt{\frac{\sum \Delta y^2}{n-1}} = 9.47 \frac{\text{МДж}}{\text{м}^2}.$$

Заключение о достоверности r можно получить непосредственно по таблице 2 путем интерполяции для уровней корреляции при $0.85 < r < 0.90$.

Таблица 2

Достаточные размеры выборок для получения достоверных значений линейного коэффициента корреляции r при доверительной вероятности $\beta = 0.95$ и $\beta = 0.99$

Величина r	Объем выборки при		Величина r	Объем выборки при	
	$\beta = 0.95$	$\beta = 0.99$		$\beta = 0.95$	$\beta = 0.99$
0.01	38407	66503	0.40	24	40
0.10	383	661	0.45	19	31
0.15	171	295	0.50	16	25
0.16	151	259	0.55	13	20
0.17	133	228	0.60	11	17
0.18	119	204	0.65	9	14
0.19	107	183	0.70	8	12
0.20	97	165	0.75	7	10
0.25	62	105	0.80	6	9
0.30	43	73	0.85	5	7
0.35	32	53	0.90	5	6

Предварительно составив таблицу 3, рассчитаем корреляционное соотношение $\eta_{x/y}$ по формуле:

$$\eta_{x/y} = \sqrt{\frac{\sum(x - \bar{x})^2 - \sum(x - \bar{x}_y)^2}{\sum(x - \bar{x})^2}} = \sqrt{\frac{350.83 - 6.50}{350.83}} = 0.99.$$

Корреляционное соотношение определяет наличие функциональной связи между переменными. В случае линейной связи оно равно квадрату значения коэффициента корреляции.

Расчет корреляционного соотношения $\eta_{x/y}$

x – значение календарных суток, когда достигается максимум численности фитопланктона за каждый год на ст. Пестово,

y – значение фотосинтетически активной солнечной радиации за февраль,

n – число родственных признаков фактора x

Год	x , сут.	y , МДж/м ²	n_y	$\bar{x}_y = \frac{\sum x}{n_y}$	$x - \bar{x}$	$(x - \bar{x})^2$	$x - \bar{x}_y$	$(x - \bar{x}_y)^2$
1998	131	66.17			8.83	78.03	0	0
1999	131	64.83	2	131	8.83	78.03	0	0
2000	114	56.67			-8.17	66.69	1.50	2.25
2002	111	41.71	2	112.50	-11.17	124.69	-1.50	2.25
2001	122	62.16			-0.17	0.03	-1	1
2003	124	66.17	2	123	1.83	3.36	1	1
Средняя	122.17	59.62				350.83		6.50

Оценим ошибку и критерий достоверности σ_η , t_η по соотношениям:

$$\sigma_\eta = \sqrt{\frac{1 - \eta^2}{n - 2}}; \quad t_\eta = \frac{\eta}{\sigma_\eta}$$

$$\sigma_{\eta_{x/y}} = \sqrt{\frac{1 - 0.98}{4}} = 0.07; \quad t_\eta = \frac{0.99}{0.07} = 14.14.$$

Эмпирический критерий t_η существенно больше табличных по Стьюденту (табл. 4) при числе степеней свободы $l = n - 2 = 4$ для уровней доверительной вероятности $t_{\beta=0.95} = 2,78$, $t_{\beta=0.99} = 4,60$ и $t_{\beta=0.999} = 8,61$ ($t_\eta > t_{\beta=0.999} > t_{\beta=0.99} > t_{\beta=0.95}$).

Таблица 4

Значения критерия достоверности по Стьюденту-Фишеру при трех уровнях вероятности P и различных числах степеней свободы (l)

Число степеней свободы, l	Уровень вероятности, P			Число степеней свободы, l	Уровень вероятности, P		
	0.95	0.99	0.999		0.95	0.99	0.999
	значения t				значения t		
1	12.71	63.66	637	19	2.09	2.86	3.88
2	4.30	9.92	31.60	20	2.09	2.85	3.85
3	3.18	5.84	12.94	21	2.08	2.83	3.82
4	2.78	4.60	8.61	22	2.07	2.82	3.79
5	2.57	4.03	6.87	23	2.07	2.81	3.77
6	2.45	3.71	5.96	24	2.06	2.80	3.75
7	2.37	3.50	5.41	25	2.06	2.79	3.73
8	2.31	3.36	5.04	26	2.06	2.78	3.71
9	2.26	3.25	4.78	27	2.05	2.77	3.69

10	2.23	3.17	4.59	28	2.05	2.76	3.67
11	2.20	3.11	4.44	29	2.05	2.76	3.66
12	2.18	3.05	4.32	30	2.04	2.75	3.65
13	2.16	3.01	4.22	35 – 39	2.03	2.72	3.59
14	2.14	2.98	4.14	40 – 44	2.02	2.70	3.55
15	2.13	2.95	4.07	45 – 60	2.01	2.66	3.50
16	2.12	2.92	4.01	70 – 100	1.98	2.63	3.39
17	2.11	2.90	3.97	120	1.96	2.58	3.29
18	2.10	2.88	3.92	>120	1.96	2.58	3.29

По полученным оценкам $r_{x/y}$ и $\eta_{x/y}$ значение критерия линейности $L_{x/y} = \eta_{x/y}^2 - r_{x/y}^2 = 0,98 - 0,74 = 0,24 > 0,10$, т.е. превышает допустимое. Следовательно, необходима дальнейшая проверка на существенность линейности связи по критерию:

$$\sigma_L = 2 \sqrt{\frac{L}{n} \sqrt{(1 - \eta^2) - (1 - r^2) + 1}} = 0,4 \sqrt{0,02 - 0,26 + 1} = 0,348 < 2,$$

где σ_L – значение статистической ошибки. Так как $\sigma_L < 2$, можно считать, что зависимость $x = f(y)$, линейна.

К аналогичному заключению приводит и другой критерий (l):

$$l = 0,742 \sqrt{n} \sqrt{\eta^2 - r^2} = 0,87 < 2.$$

Это значит, что найденная нами связь имеет линейный характер.

Следовательно, мы можем перейти к построению прогностических уравнений регрессии. Первым делом вычислим коэффициент регрессии:

$$B_{x/y} = r \frac{\sigma_x}{\sigma_y} = 0,86 \cdot \frac{8,38}{9,47} = 0,76.$$

Далее подставим значение $B_{x/y}$ в формулу: $x - \bar{x} = B_{x/y}(y - \bar{y})$. Тогда уравнение регрессии имеет вид:

$$x = 0,76y + 77. \quad (1)$$

Определим ошибку уравнения регрессии: $\sigma_{x/y} = \pm \sigma_x \sqrt{1 - r^2} = \pm 4,3$ сут. Таким образом, среднее квадратическое отклонение полученных по уравнению регрессии значений, примерно, равно 4 сут. Данный корреляционный анализ для двух параметров дал прекрасный результат: рассматриваемый временной интервал, действительно, является самым оптимальным для описания ситуации с варьированием численности фитопланктона на верхней станции Учинского водохранилища за каждый год и построения прогностического уравнения.

Теперь выполним такой же полный анализ зависимости между значениями ФАР и датами пиков «цветения» фитопланктона для ст. Листвянка. В этом случае попробуем использовать и другие методы для исследования.

Таблица 5

Вычисление линейного коэффициента корреляции r

x – значение календарных суток, когда достигается максимум численности фитопланктона за каждый год на ст. Листвянка,

y – значение интегральной фотосинтетически активной солнечной радиации за февраль

б

Год	x , сут.	y , МДж/м ²	Δx	Δy	Δx^2	Δy^2	$\Delta x \times \Delta y$
1998	131	66.17	9.50	6.55	90.25	42.90	62.26
1999	124	64.83	2.50	5.21	6.25	27.14	13.02
2000	123	56.67	1.50	-2.95	2.25	8.70	-4.42
2001	122	62.16	0.50	2.54	0.25	6.45	1.27
2002	98	41.71	-23.50	-17.91	552.25	320.77	420.90
2003	131	66.17	9.50	6.55	90.25	42.90	62.26
Σ	729	357.69			741.50	448.87	555.31
n	6	6					
Среднее	121.50	59.62					

На основе данных таблицы 5 вычислим коэффициент корреляции (r) по формуле

$$r_{x/y} = \frac{\Sigma \Delta x \Delta y}{\sqrt{\Sigma \Delta x^2 \Sigma \Delta y^2}} = \frac{555,31}{\sqrt{741,50 \cdot 448,87}} = \frac{555,31}{576,92} = 0,96.$$

Вычислим также среднеквадратические отклонения:

$$\sigma_x = \sqrt{\frac{\Sigma \Delta x^2}{n - 1}} = 12,18 \text{ сут.},$$

$$\sigma_y = \sqrt{\frac{\Sigma \Delta y^2}{n - 1}} = 9,47 \frac{\text{МДж}}{\text{м}^2}.$$

Для определения достоверности r вычислим ошибку коэффициента корреляции ($n < 50$, следовательно, по методу Z):

$$\sigma_z = \frac{1}{\sqrt{n - 3}} = 0,58,$$

где σ_z – статистическая ошибка, которая зависит только от объема выборки.

По таблице 6 находим, что $r = 0,96$ соответствует $z = 1,946$ с ошибкой $\sigma_z \pm z = 1,946 \pm 0,58$.

Таблица 6

Соотношение между величинами r и z

Десятые доли r	Сотые доли r									
	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
	значения z									
0.0	0.000	0.010	0.020	0.030	0.040	0.050	0.060	0.070	0.080	0.090
0.1	0.100	0.110	0.121	0.131	0.141	0.151	0.161	0.172	0.182	0.192
0.2	0.203	0.213	0.224	0.234	0.245	0.255	0.266	0.277	0.288	0.299
0.3	0.309	0.321	0.332	0.343	0.354	0.265	0.377	0.388	0.400	0.412
0.4	0.424	0.436	0.448	0.460	0.472	0.485	0.498	0.510	0.523	0.536
0.5	0.549	0.563	0.576	0.590	0.604	0.618	0.633	0.648	0.663	0.678
0.6	0.693	0.709	0.725	0.741	0.758	0.776	0.793	0.811	0.829	0.848
0.7	0.867	0.887	0.908	0.929	0.951	0.973	0.996	1.020	1.045	1.071
0.8	1.099	1.127	1.157	1.188	1.221	1.256	1.293	1.333	1.376	1.422
0.9	1.472	1.527	1.589	1.658	1.738	1.832	1.946	2.092	2.298	2.647

Находим выборочный порог достоверности

$$t_z = \frac{1.946}{0.58} = 3,36.$$

По критерию Стьюдента (табл. 4) определяем теоретическое значение t_z при $l = n - 2 = 4$ (n – объем выборки). Соответственно для доверительной вероятности 0.95; 0.99 и 0.999 получаем $t_{\text{табл}} = 2.78, 4.60, 8.61$. Полученная величина $r = 0.96$ значима при доверительной вероятности 0.95. Это свидетельствует о высокой достоверности оценки r .

Теперь оценим достоверность r путем интерполяции для уровней корреляции. Предварительно составив таблицу 7, рассчитаем корреляционное соотношение $\eta_{x/y}$ по формуле:

$$\eta_{x/y} = \sqrt{\frac{\sum(x - \bar{x})^2 - \sum(x - \bar{x}_y)^2}{\sum(x - \bar{x})^2}} = \sqrt{\frac{741,5 - 2}{741,5}} = 0,99.$$

Таблица 7

Расчет корреляционного соотношения $\eta_{x/y}$

x – значение календарных суток, когда достигается максимум численности фитопланктона за каждый год на ст. Листвянка,

y – значение фотосинтетически активной солнечной радиации за февраль,

n – число родственных признаков фактора y

Год	x , сут.	y , МДж/м ²	n_y	$\bar{x}_y = \frac{\sum x}{n_y}$	$x - \bar{x}$	$(x - \bar{x})^2$	$x - \bar{x}_y$	$(x - \bar{x}_y)^2$
1998	131	66.17			9.50	90.25	0	
2003	131	66.17	2	131	9.50	90.25	0	
1999	124	64.83			2.50	6.25	1	1
2000	123	56.67		123	1.50	2.25	0	
2001	122	62.16	3		0.50	0.25	-1	1
2002	98	41.71	1	98	-23.50	552.25	0	
Сумма	729	357.695				741.5		2

Оценим ошибку и критерий достоверности σ_{η} , t_{η} :

$$\sigma_{\eta_{x/y}} = \sqrt{\frac{1 - 0.98}{4}} = 0,07; \quad t_{\eta} = \frac{0.99}{0.07} = 14,14.$$

Эмпирический критерий t_{η} существенно больше табличных по Стьюденту (табл. 4) при числе степеней свободы $l = n - 2 = 4$ для уровней доверительной вероятности $t_{\beta=0.95} = 2.78$, $t_{\beta=0.99} = 4.60$ и $t_{\beta=0.999} = 8.61$ ($t_{\eta} > t_{\beta=0.999} > t_{\beta=0.99} > t_{\beta=0.95}$).

По полученным оценкам $r_{x/y}$ и $\eta_{x/y}$ значение критерия линейности $L_{x/y} = \eta_{x/y}^2 - r_{x/y}^2 = 0.98 - 0.92 = 0.06 < 0.10$, т.е. критерий линейности выполняется. Следовательно, дальнейшая проверка на существенность линейности связи необязательна: найденная нами связь имеет линейный характер.

Следовательно, мы можем перейти к построению прогностических уравнений регрессии. С этой целью вычислим коэффициент регрессии:

$$B_{x/y} = r \frac{\sigma_x}{\sigma_y} = 0.96 \cdot \frac{12,18}{9,47} = 1,23.$$

Далее подставим значение $B_{x/y}$ в формулу: $x - \bar{x} = B_{x/y}(y - \bar{y})$:

$$x = 1,23(y - 59.62) + 121,5.$$

Тогда уравнение регрессии имеет вид:

$$x = 1,23y + 48. \quad (2)$$

Определим ошибку уравнения регрессии: $\sigma_{x/y} = \pm \sigma_x \sqrt{1 - r^2} = \pm 3,4$ сут.

Таким образом, можно сделать вывод о том, что для нижней станции Учинского водохранилища полный корреляционный анализ дал хорошие результаты: подтвердилось, что связь между рассматриваемыми параметрами является сильной линейной связью, построили прогностические уравнения, которые с достаточно хорошей точностью способны описать ежегодные наступления даты максимальной численности микроводорослей в водоеме весной.

Теперь, когда удалось вывести прогностические уравнения регрессии, найдем с их помощью предполагаемые даты пика «цветения» фитопланктона на ст. Пестово и ст. Листвянка в разные годы (табл. 8).

Даты максимальной численности фитопланктона за период 1993-2013 гг.
 y – значения ФАР (МДж/м²); x_p , x_l – дата пика «цветения» фитопланктона на ст.
Пестово и ст. Листвянка соответственно (сут.)

Год	ФАР за февраль	$x_p = 0,76y + 77$	$x_l = 1,23y + 48$
1993	67.695	128	132
1994	72.435	132	137
1995	52.570	117	113
1996	74.200	133	139
1997	57.655	121	119
1998	66.170	127	129
1999	64.825	126	128
2000	56.670	120	118
2001	62.155	109	124
2002	41.705	124	99
2003	66.170	127	129
2004	60.155	123	122
2005	68.855	129	133
2006	72.745	132	137
2007	73.980	133	139
2008	48.875	114	108
2009	57.175	120	118
2010	66.585	128	130
2011	68.150	129	132
2012	66.370	127	130
2013	59.305	122	121

В таблице 8 приведены данные за двадцатилетний период интегральных значений февральской ФАР и даты максимальной численности фитопланктона на каждой из двух станций Учинского водохранилища. В результате выполненных расчетов было показано, что уравнения регрессии (1) и (2) дают достаточно хорошие приближения, то есть по данным таблицы 8 можно восстановить очень точную картину наступления пиков «цветения» микроводорослей в верховье и нижней части Учинского водохранилища.

Из проделанных расчетов следует, что:

1. для двух станций Учинского водохранилища получены высокие значения коэффициента корреляции, что говорит о сильной линейной зависимости между интегральными значениями февральской ФАР и датами максимальной численности фитопланктона;

2. в процессе выполнения анализа были получены очень маленькие статистические ошибки;

3. были построены уравнения регрессии, ошибка которых в условиях рассматриваемой задачи является допустимой.

Все выше перечисленные результаты говорят о том, что для Учинского водохранилища при прогнозировании изменения численности фитопланктона весной достаточно рассматривать только один фактор – значение ФАР за февраль. Все остальные факторы на данный процесс имеют совершенно незначительное влияние. Поэтому, учитывая ошибку, которую дают уравнения регрессии, можно строить прогноз, взяв во внимание только фактор солнечной радиации.

Список литературы

- [1]. Гмурман В. Е. Теория вероятностей и математическая статистика. М.: Высшая школа, 2003. 479 с.
- [2]. Кассандрова О.Н., Лебедев В.В. Обработка результатов наблюдений. М.: Наука, 1970. 104 с.
- [3]. Корсак М.Н., Мошаров С.А., Даллакян Г.А., Белов А.Ю., Митин А.В. Динамика фитопланктона Учинского водохранилища и биогенных элементов в 1998-1999 гг. // Вестн. Моск. ун-та. Сер. Биология. 2003. № 2. С. 34-39.
- [4]. Корсак М.Н., Мошаров С.А., Даллакян Г.А., Белов А.Ю. Особенности сезонной динамики, структуры и продуктивности фитопланктона Учинского водохранилища в 1998-2001 гг. // Вестн. Моск. ун-та. Сер. Биология. 2005. № 1. С. 33-38.
- [5]. Корсак М.Н., Мошаров С.А., Юсупова К.О., Кроленко М.И. Прогноз весеннего цветения фитопланктона в Учинском водохранилище. // Безопасность в техносфере. 2015. № 1. С. 12-17.
- [6]. Корсак М.Н., Юсупова К.О. Ретроспективный анализ влияния фотосинтетически активной радиации на весеннее «цветение» фитопланктона в водохранилище питьевого назначения. // Молодежный научно-технический вестник. МГТУ им. Н.Э. Баумана. Электрон. журн. 2015. № 8. Режим доступа: <http://sntbul.bmstu.ru/doc/799669.html> (дата обращения: 13.12.2015).