

Регионализация и регрессионный анализ температуры воздуха и осадков в глобальной БД по климату

В.Г. Коновалов, В.В. Мацковский

Институт географии РАН
119017 Москва, Старомонетный пер., 29
E-mail: vladgeo@gmail.com

Получены оценки информативности климатических факторов речного стока (осадки, температура воздуха) на различном уровне пространственного разрешения. Локальный уровень представлен многолетними наблюдениями на метеостанциях, региональный – рядами метеорологических характеристик в глобальной базе по климату CRU TS 3.0 (<http://www.cru.uea.ac.uk/cru/data/hrg-interim/>). Ряды метеоданных в глобальной базе относятся к периоду 1901-2006 гг., пространственный интервал между узлами регулярной сетки равен 0,5 градуса по долготе и широте, данные распределены по высотным уровням от 0 до 5734 м над уровнем моря. Масштаб пространственного интегрирования полей метеоданных в базе данных CRU TS 3.0 примерно соответствует полосе захвата спутниками среднего разрешения как, например, ЛАНДСАТ 7 и ТЕРРА. Учитывая возможность непосредственного использования данных этих спутников для сезонных и годовых гидрологических прогнозов, выполнен сравнительный анализ качества уравнений множественной линейной регрессии $Run=f(A_1 \dots A_N)$ и $R=f(B_1 \dots B_N)$ на примере рек Амудары и Сырдарьи – основных источников водоснабжения в бассейне Аральского моря. Здесь: Run – объем стока или расход воды, $A_1 \dots A_N$ – прямые измерения осадков и температуры воздуха, $B_1 \dots B_N$ – эти же характеристики, выбранные из базы данных CRU TS 3.0. Определение наиболее информативного состава климатических факторов стока выполняется методом полного перебора $2 \dots N$ сочетаний независимых аргументов в обучающей выборке. Информационный потенциал средней годовой температуры воздуха на региональном уровне открывает возможности для применения метода статистического даунскейлинга к выходным результатам моделей глобальной циркуляции атмосферы и океана.

Ключевые слова: температура воздуха и осадки, глобальная база данных, регионализация, сток рек Амудары и Сырдарьи, множественная линейная регрессия.

1. Цель исследования

Численное описание прошлых, нынешних и будущих изменений речного стока на региональном уровне является предметом интереса для научной гидрологии, управления водными ресурсами и водного хозяйства. В данном случае понятие речного стока относится к естественному (неискаженному) потоку, который может быть измерен на нижней границе зоны формирования стока. Это означает, что процесс поверхностного стока является функцией только климатических факторов. Поскольку поверхностный сток является интегральной суммой входных воздействий, его климатические факторы должны быть выражены в виде полей, например, осадков, температуры воздуха и т. д. Однако, вместо оценки полей, как правило, используется концепция репрезентативных или информативных пунктов с целью изучения и использования соотношений между стоком и его климатическими факторами. Этот пункт может быть метеорологической станцией или узлом регулярной сетки с известными характеристиками климата и заданными географическими координатами и высотой над уровнем моря. Априорно первый вариант является более надежным, поскольку он основан на прямых измерениях влияющих переменных. Однако, для оценки временной изменчивости стока, второй вариант дает возможность использовать данные климатических архивов с высоким пространственным разрешением (Hijmans et al., 2005; Mitchell et al., 2004; Yatagai et al., 2008). По сравнению с использованием

метеоданных на локальном уровне (Агальцева, 2002), более перспективным для гидрологических расчетов представляется построение корреляционных связей между температурой воздуха и осадками в узлах регулярной сетки высокого разрешения и результатами расчета пространственного распределения этих же характеристик с помощью существующих моделей общей циркуляции атмосферы и океана (см., например, Wigley, 2008).

Таким образом, наша цель состоит в изучении и проверке информационного потенциала обеих вариантов климатических факторов на примере двух крупнейших рек внутри бассейна Аральского моря.

2. Область исследования и общая информация

Изучаемый регион представляет собой замкнутый (бессточный) бассейн Аральского моря и включает в себя водосборы рек Амударья, Сырдарья, Теджен, и Мургаб. Реки Сырдарья и Амударья являются основными источниками притока воды в Аральское море. Площадь бассейнов этих рек равна соответственно 692,3 и 493,0 тыс. км². Область формирования стока составляет около 25% площади бассейна Аральского моря. При этом 80% этой площади расположено в бассейнах рек Амударья и Сырдарья. Их зоны формирования стока могут быть приблизительно оценены как 200 000 км² для Амударии и 160 000 км² для Сырдарьи (см. рис. 1). Годовой и сезонный сток в бассейнах рек Амударья и Сырдарья подвержен многолетним изменениям. В экстремально многоводном 1969 году годовой сток в бассейне реки Амударья был равен почти 110 км³, а в маловодном 1974 году сток составил около 65 км³. В бассейне реки Сырдарьи в том же 1969 году сток за год был равен 70 км³, а в маловодным 1983 году был около 20 км³.

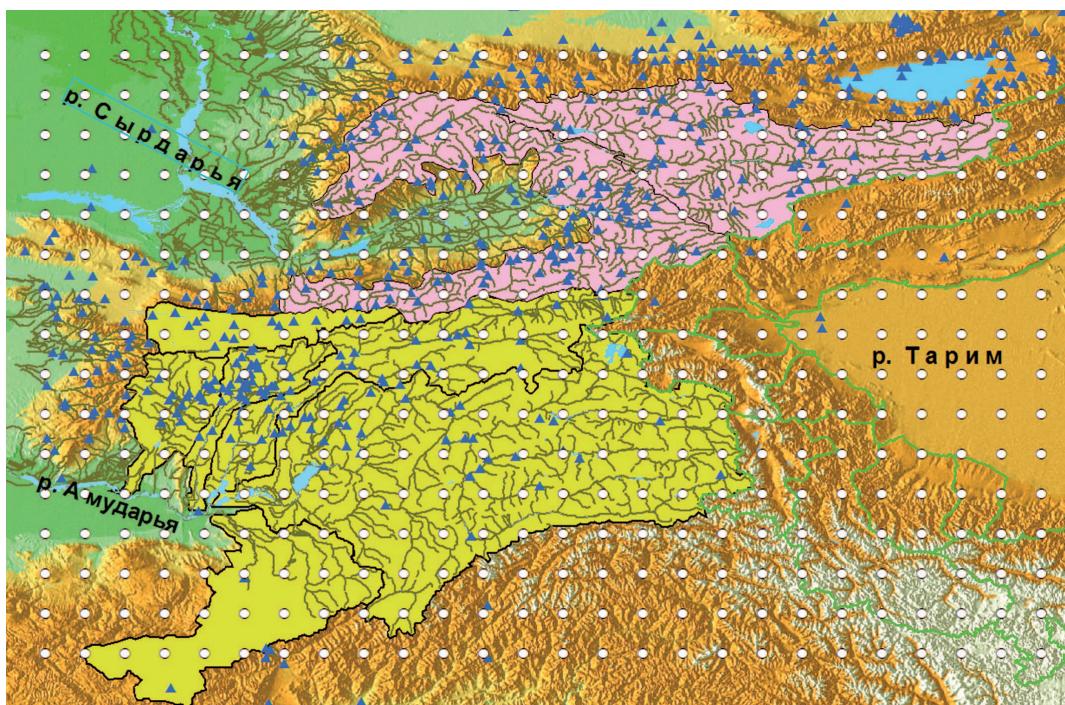


Рис 1. Бассейны рек Амудары (заливка желтым цветом) и Сырдары (заливка розовым цветом) и их притоков. Белые кружки – данные по температуре воздуха (T) и осадкам (P) в сеточном климатическом архиве CRU TS 3.0, синие треугольники – измерения T и P на метеостанциях

Исходные данные

В качестве климатических факторов речного стока (Run) использованы сумма осадков (P) за январь-декабрь и средняя за год температура воздуха (T), представленные тремя вариантами данных. Первый (локальный уровень) – это измерения на 27 длиннорядных метеорологических станциях, расположенных в интервалах: по широте $37,60^{\circ}$ - $43,23^{\circ}$ с.ш., долготе $62,18^{\circ}$ - $78,20^{\circ}$ в.д., и высоте 193-4 169 метров над уровнем моря. Источники данных: (Former Soviet Union Monthly Precipitation Archive, 1891-1993; Vose et al., 1992; Williams and Konovalov, 2008). Второй вариант (региональный уровень) – значения P и T , с шагом 0,5 градуса по широте и долготе, извлеченные для региона из Глобального климатического архива CRU TS 3.0 (Mitchell et al., 2004). Пространственные границы для этих данных по долготе $65,0^{\circ}$ - $78,0^{\circ}$ в.д., широте $34,0^{\circ}$ - $42,0^{\circ}$ с.ш. и высоте 2000-3000 м над уровнем моря Третий вариант, относящийся к данным метеорологических станций и сеточным значениям P и T , был так называемый индекс (Ib) баланса (Коновалов, 1985, 2006), определенный в каждом году, как разность безразмерных аномалий P и T : $Ib=I_p-I_T$, где $I_p=(Pi-meanP)/meanP$, а $I_T=(Ti-meanT)/meanT$. Как было показано ранее (Коновалов, 1985, 2006), этот индекс обладает хорошей информативностью и используется для повышения качества гидрологических и гляциологических расчетов. Основными источниками исходных данных по стоку были (Bodo, 2000) и (Основные гидрологические характеристики ..., 1967-1980). Интервал времени, равный базовому климатическому периоду 1961-1990 гг, был выбран для описания изменчивости стока рек, как функции осадков и температуры воздуха.

3. Методы

Множественная линейная регрессия третьего порядка принята в качестве общей формы связи между стоком и его климатическими факторами. Уравнение регрессии $Run=f(P,T)$ между стоком и его климатическими факторами можно считать неявным и упрощенным представлением уравнения годового водного баланса речного бассейна:

$$Run = K_R(P - E) \pm \Delta W, \quad (1)$$

поскольку осадки входят в оба уравнения, а температура воздуха в уравнении регрессии более или менее удачно служит косвенной характеристикой остатка $Run-P$. В уравнении (1) K_R – коэффициент трансформации поступления влаги на поверхность бассейна, E – испарение, ΔW – динамические запасы воды в бассейне. Выявление информативных пунктов для описания пространственно-временной изменчивости климатических факторов стока принято в качестве метода регионализации исходных данных по осадкам и температуре воздуха.

В соответствии с поставленной задачей, значения годового стока Амударьи, Сырдарьи и суммы для обеих рек использованы в качестве предиктантов или зависимых переменных в уравнениях: $Run=f(A_1 \dots A_N)$ и $Run=f(B_1 \dots B_N)$. Здесь: Run – объем поверхностного стока или расход воды, $A_1 \dots A_N$ – прямые измерения осадков или температуры воздуха на метеостанциях, $B_1 \dots B_N$ – эти же характеристики, выбранные из базы данных CRU TS 3.0.

Для анализа полного набора возможных комбинаций предикторов в эмпирических уравнениях множественной линейной регрессии и поиска наивысшего коэффициента кор-

реляции использован специальный алгоритм. На первом этапе расчета в программе регрессии (см. рис 2) выявляется наивысший коэффициент корреляции для уравнения первого порядка: предиктант = f (предиктор). Далее рассматриваются все возможные сочетания предикторов в уравнении второго порядка и выбирается такая их комбинация, которая дает наибольший прирост коэффициента корреляции, полученного на первом этапе расчета. Аналогичная процедура выполняется для получения наилучшего уравнения третьего порядка. На каждом этапе определяются следующие параметры регрессии.

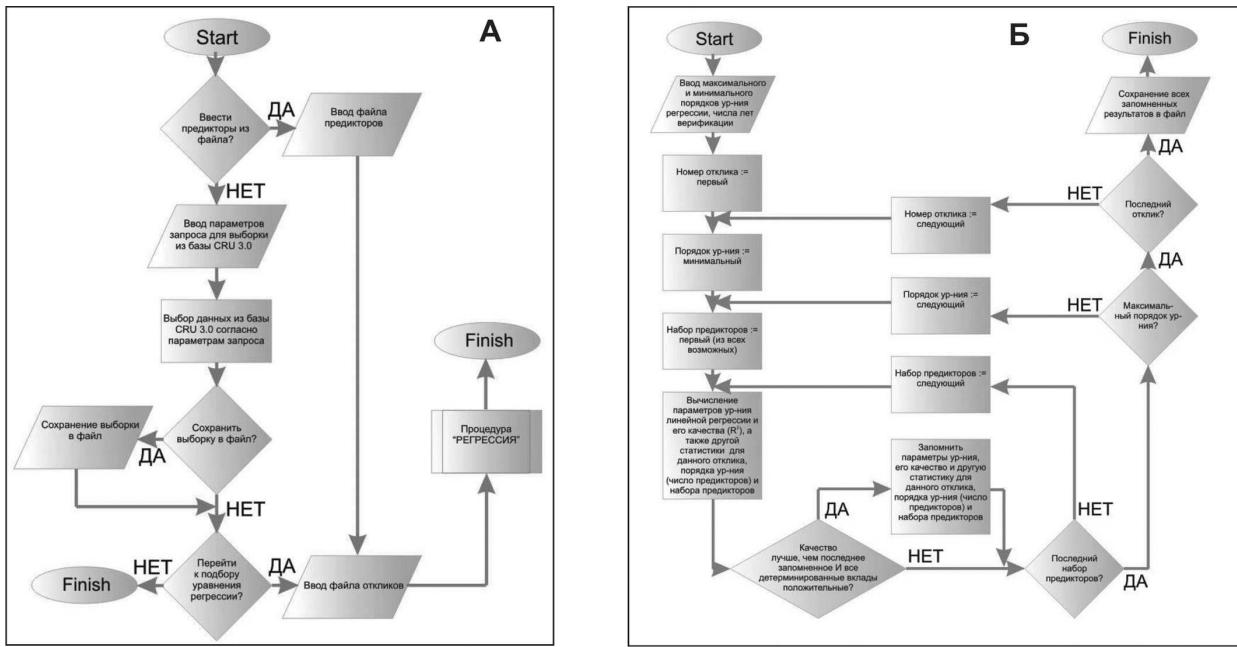


Рис. 2. а – основной алгоритм работы программы множественной регрессии,
б – алгоритм процедуры «Регрессия»

P_{ur} – порядок уравнения регрессии; R_c – сводный коэффициент корреляции для уравнения регрессии порядка P_{ur} ; S_r – ошибка определения R_c ,

$$S_r = \frac{1 - R_c^2}{\sqrt{N - P_{ur}}} \quad (2)$$

$S_y = \sigma_y \sqrt{1 - R_c^2}$ – ошибка расчета предиктанта по уравнению регрессии, S_y/σ_y – оценка качества уравнения регрессии, R_{inf} – минимальная граница доверительного интервала для коэффициента корреляции с принятой вероятностью t_p , R_{sup} – максимальная граница доверительного интервала для коэффициента корреляции с принятой вероятностью t_p , $R_{inf} < R^* < R_{sup}$;

$$z_1 = z - \sigma_z t_p < z^* < z + \sigma_z t_p = z_2;$$

$$\sigma_z = \frac{1}{\sqrt{N - P_{ur} - 2}} \quad (3)$$

N – длина временного ряда, $t_p = 1,96$;

$$z = 0.5 \cdot \ln \frac{1 + R_c}{1 - R_c} \quad (4)$$

$R_{inf} = thz_1; R_{sup} = thz_2$; th – гиперболический тангенс; Nx – номер предиктора, выбранного в уравнение регрессии порядка P_w ; β_j – детерминированный вклад j -го предиктора в уравнение нормированной регрессии, сумма детерминированных вкладов должна быть равна единице. Учитывая заданный размер обучающей выборки ($N = 30$ лет), максимальный порядок уравнений регрессии принят равным трем.

Детерминированные вклады предикторов в описание дисперсии функции были рассчитаны по формуле Г.А. Алексеева (1971), которая в частности для трех предикторов имеет вид:

$$\beta_j = \frac{|r_{0j}\alpha_{0j}|}{|r_{01}\alpha_{01}| + |r_{02}\alpha_{02}| + |r_{03}\alpha_{03}|} \quad (5)$$

$$\beta_1 + \beta_2 + \beta_3 = 1$$

Здесь r_{0j} – парные коэффициенты корреляции предиктанта с j -м предиктором, α_{0j} – коэффициенты уравнения нормированной регрессии.

Окончательные результаты поиска наиболее эффективных уравнений типа $Run = f(T)$, $Run = f(P)$ и $Run = f(Ib)$ представлены в таблице 1. Таблица 2 характеризует результаты проверки полученных уравнений по данным контрольной выборки за 1932-1961 гг.

Таблица 1. Характеристики уравнений регрессии 3-го порядка для расчета годового стока.

Получены на основе выборок за 1961-1990

Бассейн		Данные метеостанций					Данные 82 узлов в БД CRU TS 3.0				
		T		P		Ib		T		P	
Амударья		R	0,80	0,88		0,91		0,72		0,87	
		δR	0,06	0,04		0,03		0,09		0,05	
		$Crit$	0,58	0,47		0,42		0,69		0,49	
	Cont	мст 5	0,35	мст 1	0,25	мст 1	0,24	g 37	0,05	g 52	0,87
		мст 6	0,41	мст 3	0,31	мст 6	0,65	g 68	0,33	g 69	0,02
		мст 8	0,24	мст 6	0,44	мст 9	0,11	g 79	0,62	g 76	0,10
Амударья + Сырдарья		R	0,73	0,86		0,88		0,68		0,83	
		δR	0,06	0,05		0,04		0,10		0,06	
		$Crit$	0,69	0,51		0,48		0,73		0,55	
	Cont	мст 9	0,35	мст 2	0,15	мст 3	0,56	g 37	0,30	g 18	0,56
		мст 10	0,58	мст 3	0,59	мст 4	0,40	g 66	0,50	g 42	0,06
		мст 8	0,07	мст 4	0,27	мст 8	0,04	g 72	0,20	g 44	0,38
Сырдарья		R	0,61	0,89		0,89		0,73		0,87	
		δR	0,08	0,04		0,04		0,09		0,05	
		$Crit$	0,80	0,46		0,45		0,68		0,49	
	Cont	мст 1	0,52	мст 1	0,23	мст 1	0,26	g 64	0,17	g 50	0,79
		мст 5	0,13	мст 3	0,47	мст 3	0,34	g 73	0,10	g 72	0,08
		мст 7	0,34	мст 6	0,30	мст 6	0,40	g 75	0,73	g 76	0,13

Примечание: R – коэффициент корреляции; δR – ошибка расчета R ; $Crit$ – критерий качества эмпирического уравнения (равен отношению Sy/σ_y , где Sy ошибка расчета зависимой переменной Y , а σ_y стандарт отклонения Y); мст – метеорологическая станция, данные которой используются как независимая переменная; Cont – детерминированный вклад независимой переменной в описание стандарта отклонения Y ; $Ib = IP - IT$ – индекс баланса; IP – аномалия годовой суммы осадков в i -ом году, деленная на среднее значение временного ряда; IT – аномалия средней годовой температуры воздуха в i -ом году деленная на среднее значение временного ряда; g – узел в базе данных CRU TS 3.0. Расшифровка названий метеостанций: мст 1 – Алматы; мст 2 – Дехавз; мст 3 – Джизак; мст 4 – Ош; мст 5 – Пскем; мст 6 – Ташкент; мст 7 – Тянь-Шань; мст 8 – ур. Тос; мст 9 – ур. Андижан; мст 10 – Ленинабад

4. Результаты и выводы

4.1. В целом ежегодные значения осадков и температуры воздуха являются достаточно информативными переменными для получения линейных уравнений $\text{Run} = f(T)$, $\text{Run} = f(P)$ и $\text{Run} = f(Ib)$, обеспечивающих приемлемое качество гидрологических расчетов.

4.2. Оба варианта данных по осадкам (то есть, измеренных на метеостанциях и рассчитанных в узлах сетки) оказались лучшими предикторами, чем температура воздуха. Уравнения $\text{Run} = f(P)$ также показывают лучшие результаты при проверке на независимых данных (таблица 2).

4.3. Набор предикторов выбранных для правой части уравнений $\text{Run} = f(T)$, $\text{Run} = f(P)$ и $\text{Run} = f(Ib)$, как правило, включает одни и те же пункты измерений, что подтверждает надежность статистических связей.

4.4. Результаты применения индекса баланса в качестве предиктора характеризуют его более высокую информативность и эффективность, в сравнении с температурой воздуха и осадками.

4.5. Уравнения $\text{Run} = f(T)$, полученные по данным о температуре воздуха в узлах регулярной сетки с разрешением 0,5 градуса по широте и долготе, могут быть использованы совместно с другими подходами для сверхдолгосрочных прогнозов годового стока Амударьи, Сырдарьи и Амударьи + Сырдарья, когда будущие значения предикторов определяются путем даунскейлинга на региональный или локальный уровень выходных результатов моделей общей циркуляции атмосферы и океана.

Работа выполнена при поддержке гранта РФФИ 08-05-00661.

Таблица 2. Качество расчета годового стока по уравнениям регрессии 3-го порядка в обучающей (1961-90 гг.) и контрольной выборках (1932-60 гг.)

Бассейн	Измерения независимых переменных на метеостанциях					
	Температура воздуха Т		Осадки Р		Индекс баланса Ib	
	1961-1990	1932-1961	1961-1990	1932-1960	1961-1990	1932-1960
Амударья	11,62	16,94	5,12	6,87	4,17	12,62
Амударья + Сырдарья	10,28	17,53	7,21	10,03	7,95	11,31
Сырдарья	11,33	12,22	7,56	12,14	7,14	12,52
Информация из базы данных CRU 3.0						
Амударья	7,06	12,30	5,22	7,49	4,17	7,49
Амударья + Сырдарья	10,47	10,12	7,89	13,16	7,14	12,63
Сырдарья	9,78	16,82	7,13	10,51	6,44	10,62

Литература

1. Агальцева Н.А. Оценка влияния климатических изменений на располагаемые водные ресурсы в бассейне Аральского моря // Диалог о воде и климате. НИЦ МКВК, Ташкент, 2002, С. 3-59.
2. Алексеев Г.А. Объективные методы выравнивания и нормализации корреляционных связей. Л.: Гидрометеоиздат, 1971. 362 с.
3. Коновалов В.Г. Таяние и сток с ледниками в бассейнах рек Средней Азии // Л.: Гидрометеоиздат, 1985. 236 с.
4. Коновалов В.Г., Вильямс М.В. Многолетние колебания оледенения и стока рек Центральной Азии в современных климатических условиях // Метеорология и гидрология, 2005. № 9. С. 69–83.

5. Коновалов В.Г. Региональная модель гидрологического режима ледников (РЕГМОД) // Оледенение Северной и Центральной Евразии в современную эпоху: Т. 1. М.: Наука, 2006. С. 338-380.
6. Основные гидрологические характеристики и государственный Водный Кадастр СССР, 1967-1980. Том 14, вып. 1. Бассейн р. Сырдарьи. – Л.: Гидрометеоиздат, 1967. 477 с.; 1974. 529 с.; Том 14, вып. 3. Бассейн р. Амударьи. – Л.: Гидрометеоиздат, 1976. 434 с.; 1980. 443 с.
7. Bodo, B. A. Monthly Discharges for 2400 Rivers and Streams of the former Soviet Union (FSU). Toronto, Canada, 2000. V. 1.0.
8. Former Soviet Union Monthly Precipitation Archive, 1891-1993 // NSIDC-0059. 1998. Boulder, USA. <http://www-nsidc.colorado.edu/>.
9. Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., and Jarvis A. Very high resolution interpolated climate surfaces for global land areas. // Int. J. Climatol., 2005. 25. P. 1965–1978.
10. Mitchell, T.D., Carter, T.R., Jones, P.D., Hulme, M., New, M. A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1901-2000) and 16 scenarios (2001-2100) // Tyndall Working Paper, 2004. No. 55. Tyndall Centre, Norwich, UK, 30 p.
11. Vose, R. S., R. L. Schmoyer, P. M. Steurer, T. C. Peterson, R. Heim, T. R. Karl, and J. Eischeid. The Global Historical Climatology Network: Long-term monthly temperature, precipitation, sea level pressure, and station pressure data // ORNL/CDIAC-53 NDP-041, 1992. 324 p.
12. <http://www.ncdc.noaa.gov/ghcnm/>
13. Wigley, T.M. MAGICC/SCENGEN 5.3: USER MANUAL (version 2) // NCAR, 2008. Boulder CO, 81 p.
14. Williams, M.W. and Konovalov, V.G. Central Asia Temperature and Precipitation Data, 1879-2003 // Boulder, Colorado, USA, 2008. National Snow and Ice Data Center. http://nsidc.org/data/docs/noaa/g02174_central_asia_data/index.html
15. Yatagai, A., Xie, P., Alpert, P. Development of a daily gridded precipitation data set for the Middle East. // Adv. Geosci., 2008.12. P. 165–170.

Regionalization and regression analysis of air temperature and precipitation in global Data Base on climate

V.G. Konovalov, V.V. Matskovskiy

Institute of Geography RAS
29, Staromonetny per., 119017, Moscow
E-mail: vladgeo@gmail.com

Estimates of informative capacity were obtained for the climatic factors of river runoff (precipitation, air temperature) at different levels of spatial resolution. The local level is presented by long-term observations at meteorological stations, regional one – by series of meteorological characteristics in the global DB on climate CRU TS 3.0 (<http://www.cru.uea.ac.uk/cru/data/hrg-interim/>). Series of meteorological data in this global data base refers to the period of 1901-2006 years. Spatial interval between the nodes of a regular grid is equal to 0,5 degrees in longitude and latitude. Data are distributed along altitude from 0 to 5,734 m above sea level. The scale of spatial integration of fields of meteorological data in the database CRU TS 3.0 is roughly equivalent to the swath area of satellites of medium resolution such as LANDSAT 7 and TERRA. Given the possibility of direct use data from these satellites to seasonal and annual hydrological forecasts, the comparative analysis was performed for the quality of multiple linear regression equation $\text{Run} = f(A_1 \dots A_N)$ and $R = f(B_1 \dots B_N)$ on an example of the Amu Darya and Syr Darya rivers – the main sources of water supply in Aral Sea Basin. Here: Run – runoff or water flow, A₁ ... A_N – direct measurements of precipitation and air temperature, B₁ ... B_N – the same characteristics, selected from a database CRU TS 3.0. Determining the most informative composition of the climatic factors of runoff is performed by the method of exhaustive search 2 ... N combinations of independent variables in the training sample. Information capacity of the average annual air temperature at the regional level offers an opportunity to apply the method of statistical downscaling to the output results of models of global circulation of atmosphere and ocean.

Keywords: air temperature and precipitation, global database, regionalization, runoff of the Amu Darya and Syr Darya rivers, multiple linear regression.