

Н.И. Сенцова

ГИДРОЛОГИЧЕСКОЕ РАЙОНИРОВАНИЕ РЕЧНЫХ БАССЕЙНОВ ПО СЕЗОННОЙ ИЗМЕНЧИВОСТИ ХАРАКТЕРИСТИК РЕЧНОГО СТОКА¹

Рассматривается районирование речных бассейнов по характеристикам сезонной изменчивости речного стока. Для целей классификации на основе анализа временных рядов среднемесячных расходов воды сгенерированы новые ряды характеристик речного стока, учитывающих сезонную изменчивость. В результате многовариантных экспериментов в качестве наиболее информативных показателей внутригодовых изменений речного стока были выделены такие характеристики, как индексы сезонности, корреляционные функции и функции спектральных плотностей.

Для выявления групп однородных объектов использовалась программная система «TeleStat», предназначенная для статистического анализа многомерных данных и временных процессов. Районирование производилось с помощью кластерного анализа с использованием регрессионной модели на основе радиальных базисных функций. В качестве переменных использовались полученные наборы численных показателей для каждого пункта наблюдений. Веса переменных подбирались так, чтобы обеспечить сочетание как географической (координатной), так и критериальной близости объектов (гидрологических постов наблюдений). Проведенные процедуры регионализации по характеру сезонных колебаний речного стока позволили выделить однородные гидрологические районы на территории ряда крупных бассейнов рек России, в частности Волги и верхнего Дона.

Классификация рек дала возможность провести совместный анализ по группам однородных гидрологических объектов. Для каждого гидрологического района рассчитаны обобщенные статистические характеристики колебаний речного стока. Такие статистические параметры речного стока могут служить для его дальнейшего моделирования.

Применение расчетных и прогнозных методов по наблюдениям на отдельных гидрологических постах часто затруднено по разным причинам: неравномерное распределение пунктов наблюдений, недостаточная продолжительность временных рядов для статистического анализа, наличие трудно восстанавливаемых пропусков, значительных погрешностей в измерениях и т.д. Поэтому, несмотря на высокую компьютеризацию и наличие современных программных паке-

¹ Работа выполнена при финансовой поддержке РФФИ, грант №08-05-00807-а.

тов, возникают сложности, связанные как с обработкой больших массивов гидрологических данных, так и с их последующим использованием в моделях. Необходимо обрабатывать и анализировать каждый ряд, восстанавливать пропуски, искать выбросы, возможные ошибки и пр. С целью увеличения информации и для выявления возможных случайных ошибок следует проводить совместный анализ по группам однородных гидрологических объектов. Обобщенные статистические параметры речного стока, полученные для таких групп, могут служить для его дальнейшего моделирования.

В связи с этим следует проводить анализ временных рядов среднемесячных расходов воды с последующим выделением однородных гидрологических районов и получением для каждого из них обобщенных (региональных) характеристик колебаний речного стока. Другими словами, должна быть разработана их классификация.

Вопросам классификации рек и гидрологическому районированию уделяется много внимания. Классификация рек подразумевает их распределение по группам на основании выявленных наиболее существенных признаков, а районирование – результат обобщения полученной классификации для конкретной территории, что позволяет определять анализируемые характеристики слабоизученных объектов по аналогии с изученными. При решении задач, связанных с оценкой пространственного распределения речного стока и его сезонной изменчивости в современных условиях, возникает необходимость разработки новых вариантов районирования.

Выполненные нами научные исследования позволяют провести гидрологическое районирование ряда крупных речных бассейнов на территории РФ. С помощью разработанных стохастических моделей и методов оценивания были изучены особенности сезонных колебаний речного стока в различных физико-географических условиях (Болгов и др., 2005).

Основой для выбора районов послужил анализ внутригодового распределения речного стока путем формализации показателей его сезонной изменчивости и объединения (группирования) временных рядов среднемесячных расходов воды по разработанным для классификации алгоритмам.

Проведена предварительная обработка многолетних наблюдений за стоком воды на реках России. Информационной основой исследований послужили данные многолетних наблюдений за стоком рек на территории России. Из существующего архива данных на-

блюдений Росгидромета на гидрологических постах, содержащего более 5000 временных рядов со среднемесячными расходами воды, была проведена выборка рядов наблюдений, которые могут быть использованы для гидрологического районирования территории. С этой целью все имеющиеся в наличии временные ряды были исследованы на однородность, длительность, наличие пропусков и выбросов. Стоковые ряды, содержащие трудно восстанавливаемые пропуски, ряды длительностью менее 10 лет, а также подверженные значительному антропогенному воздействию (в частности, созданию водохранилищ), были исключены из дальнейшего анализа. В процессе первичной обработки были выявлены и восстановлены аномальные выбросы, заполнены пропуски в наблюдениях, рассчитаны необходимые статистические характеристики для каждого временного ряда. В результате был получен массив данных более чем по 1700 гидрологическим постам, содержащий файлы со среднемесячными расходами воды и координатами центров тяжести водосбора для каждого пункта наблюдений.

Для целей классификации на основе анализа временных рядов среднемесячных расходов воды сгенерированы новые ряды характеристик речного стока, учитывающие сезонную изменчивость. В результате многовариантных экспериментов в качестве наиболее информативных показателей внутригодовых изменений речного стока были выделены такие характеристики, как индексы сезонности, корреляционные функции и функции спектральных плотностей.

Влияние внешних факторов, действующих циклически с заранее известной периодичностью, определяет сезонные колебания речного стока. Наличие сезонных эффектов проявляется в виде острых узких пиков в периодограмме на соответствующей частоте (а при несимметричной форме сезонной волны – и на кратных частотах). В выборочной автокорреляционной функции также присутствуют выбросы для лагов (запаздываний), кратных периоду сезонности, но эти выбросы могут быть завуалированы присутствием тренда или большой дисперсией случайной компоненты.

Одна из наиболее простых моделей учета сезонности – модель сезонных эффектов, она является одной из принятых формализаций сезонных колебаний речного стока. В аддитивной форме этой модели ряд представляется в виде

$$Y(t) = T(t) + S(t) + e_t,$$

где $T(t)$ – тренд, e_t – ошибка, а $S(t)$ – сезонная составляющая, которая определяется своими значениями на сезонном периоде длины L , $S(1)$, ... $S(L)$. Для однозначности параметризации модели обычно предполагают, что $S(1) + \dots + S(L) = 0$. Значения $S(1)$, ... $S(L)$ называют индексами сезонности. В настоящей работе использовались нормированные индексы сезонности, определенные на основе методов декомпозиции по аддитивной модели Тейла-Важа (Кендалл, Стюарт, 1976).

Также была исследована частотная структура сезонных колебаний речного стока. Типичная функция спектральной плотности для рассматриваемых рядов среднемесячных расходов воды содержит четко выделенные максимумы (пики) на частотах, соответствующих периодам в 12, 6, 4 и 3 месяца. Поэтому для приближенного описания функции спектральной плотности рекомендуются ее относительные значения на этих частотах. Для учета остальных составляющих спектральной плотности используется величина, рассчитанная как разность между единицей и суммой относительных значений спектральной плотности на выделенных частотах.

Наборы показателей, полученные в результате обработки временных рядов наблюдений за среднемесячными расходами воды на реках России, выявленные в результате многовариантных экспериментов с разными наборами характеристик для более точного описания сезонных колебаний речного стока, послужили для гидрологического районирования на основе разработанных стохастических моделей и статистических методов оценивания.

Для выявления групп однородных объектов использовалась программная система «TeleStat», предназначенная для статистического анализа многомерных данных и временных процессов (Енюков, Ретинская, Скуратов, 2004). В процессе исследований данная система адаптировалась для анализа гидрологических данных, с учетом специфики имеющейся исходной информации, типичной для территории РФ (неравномерности распределения пунктов наблюдений по территории, временной неоднородности рядов, различной длительности наблюдений, наличия пропусков и аномальных выбросов и др.).

Районирование производилось с помощью кластерного анализа с использованием регрессионной модели на основе радиальных базисных функций. В качестве переменных использовались полученные наборы численных показателей для каждого пункта наблюдений. Веса переменных подбирались так, чтобы обеспечить сочетание как географической (координатной), так и критериальной близости

сти объектов (гидрологических постов наблюдений). Включение координат центров тяжести водосборов как в зависимые (предикторные), так и в критериальные переменные позволило выделить однородные группы с учетом пространственного размещения гидрологических объектов.

Остановимся подробнее на процедурах регионализации с помощью кластерного анализа на основе радиальных базисных функций.

Решение проблемы одновременной группировки (кластеризации) рек по параметрам речного стока и в пространстве географических координат (районирование) основано на предположении, что географическая близость объектов влечет и близость соответствующих им характеристик речного стока (зависимых переменных). В этом случае подходящим методом является кластеризация на основе радиальных базисных функций (РБФ) в пространстве географических координат – широты S и долготы L . Далее эти переменные также именуются предикторными.

Критерий оптимальности группировки. Задача формализована следующим образом. Пусть имеется N объектов (рек), которые должны быть распределены на k групп. Число групп и их состав заранее неизвестны и должны быть определены в процессе решения задачи. При фиксированном числе классов критерием, по которому оценивается качество группировки объектов, служит остаточная дисперсия.

В качестве РБФ используется негативная экспоненциальная функция от расстояния между центром РБФ и объектом, а в качестве расстояния – взвешенное евклидово расстояние.

Модификация информационного критерия Акайке используется как критерий качества группировки с учетом числа классов.

При возрастании числа классов остаточная дисперсия и, следовательно, первое слагаемое в критерии Акайке убывает, тогда как второе слагаемое возрастает. Из двух группировок с разным числом классов более предпочтительной считается группировка с меньшим значением критерия Акайке (Енюков, Ретинская, Скуратов, 2004).

Алгоритм группировки. Алгоритм группировки основан на последовательном удалении «плохих» центров групп и состоит в следующем. Из N исходных объектов случайным образом выбирается достаточно большое количество ($M \approx 3\sqrt{N}$) объектов, которые рассматриваются как начальные центры классов. Для всех имеющихся центров пересчитываются остаточные значения для группировок, получающихся, если из текущего набора центров удаляется этот центр. Определяется центр, без которого остаточная дисперсия новой группиров-

ки увеличивается на минимальную величину. Проверяется значение критерия Акайке. Если его значение уменьшилось, из набора удаляется найденный центр и опять пересчитываются остаточные значения для группировок. В противном случае процедура останавливается. Количество групп равно количеству оставшихся центров. Объекты распределяются по группам по следующему правилу. Объект относится к той группе, до центра которой расстояние от объекта минимально (в пространстве предикторных переменных). Окончательно группировка выбирается по минимуму критерия Акайке по всем повторам.

Анализ резко выделяющихся (аномальных) значений наблюдений. Устойчивые оценки. Для получения устойчивых оценок средних, дисперсий и корреляций в исследовании производятся процедуры визуального анализа резко выделяющихся (аномальных) наблюдений. С этой целью используется подход на основе целенаправленного проецирования. Этот метод основан на целенаправленном проецировании с выбором проекций, подходящих для анализа аномальных наблюдений. Критерием, по которому ищутся такие проекции, является отношение двух оценок дисперсии – обычной и устойчивой, поскольку аномальные наблюдения существенно увеличивают величину оценки дисперсии. Таким образом, выбирается направление проецирования, на котором стандартная оценка дисперсии в наибольшей степени «испорчена» присутствием аномальных наблюдений. Хорошее приближение для определения направления, максимизирующего отношение дисперсий, получается из решения обобщенной задачи на собственные значения: $(S-lC)U=0$, где S – обычная оценка ковариационной матрицы; C – устойчивая оценка матрицы рассеивания; l – собственное число; U – направление проецирования.

Экспоненциально-взвешенная оценка используется в качестве устойчивой оценки. В качестве приближенного решения для направлений проецирования используются собственные вектора задачи, соответствующие первому, второму и т. д. по величине собственным числам. Всего же получается p векторов, упорядоченных по убыванию собственных чисел.

Таким образом, в результате применения вышеописанных процедур и методов выполняется гидрологическое районирование, отражающее особенности формирования сезонных колебаний речного стока, выявленные стохастическими методами.

Для примера рассмотрим результаты исследований по бассейнам рек Волги и верхнего Дона.

Бассейн реки Волги пересекает с севера на юг несколько физико-географических зон. Основная питающая часть водосбора Волги, от истока до Нижнего Новгорода и Казани, расположена в лесной зоне, средняя часть бассейна, до Самары и Саратова, – в лесостепной зоне, нижняя часть, до Волгограда, – в степной зоне, а южнее – в полупустынной зоне. Волгу принято делить на три части: верхняя Волга – от истока до устья Оки, средняя Волга – от впадения Оки до устья Камы и нижняя Волга – от впадения Камы до устья.

Бассейн верхнего Дона расположен в Центрально-Черноземном регионе России в пределах лесостепной зоны. С помощью предложенных выше методов и процедур выполнено районирование бассейнов Волги и верхнего Дона (рис. 1 и 2). В результате получены однородные гидрологические районы по сезонной изменчивости речного стока (Сенцова, 2008; Сенцова, 2009).

Как видно из рис. 1, в бассейне Волги четко выделяются верхне-волжская часть, бассейны рек Оки и Камы, а также средняя и нижняя Волга. Для исследования из исходной базы данных было извлечено около 90 гидрологических постов с рядами наблюдений, удовлетворяющих вышеописанным условиям регионализации. К сожалению, из-за отсутствия современных данных, ряды были ограничены 1985 г.

Бассейн верхнего Дона до г. Лиски разделился на две части – левобережную и правобережную (рис. 2). Следует отметить, что для регионализации территории были использованы данные наблюдений по действующим по настоящее время 20 постам (до 2007 г.), учитывающие влияние современных климатических изменений.

Результаты проведенных исследований сопоставлялись с результатами существующего районирования по характеру питания рек с учетом географической зональности. Для этого была произведена оцифровка районов, выделенных в работах П.С. Кузина и др. (Кузин, 1960; Кузин, Бабкин, 1979). С помощью специально разработанных компьютерных программ для каждого района производился отбор гидрологических постов наблюдений из общей базы данных. Это позволило получить карту существующего районирования в электронном виде. Затем, на основе визуального анализа и построения таблиц сопряженности, сравнивались разные результаты районирования. Таблицы сопряженности представляют собой двухфакторные (двухходовые) таблицы. Районы, выделенные в результате одного варианта районирования, образуют колонки таблицы, а районы, полученные по другому варианту – строки. Выбор «наилучшего» варианта классификации осуществляется по величи-

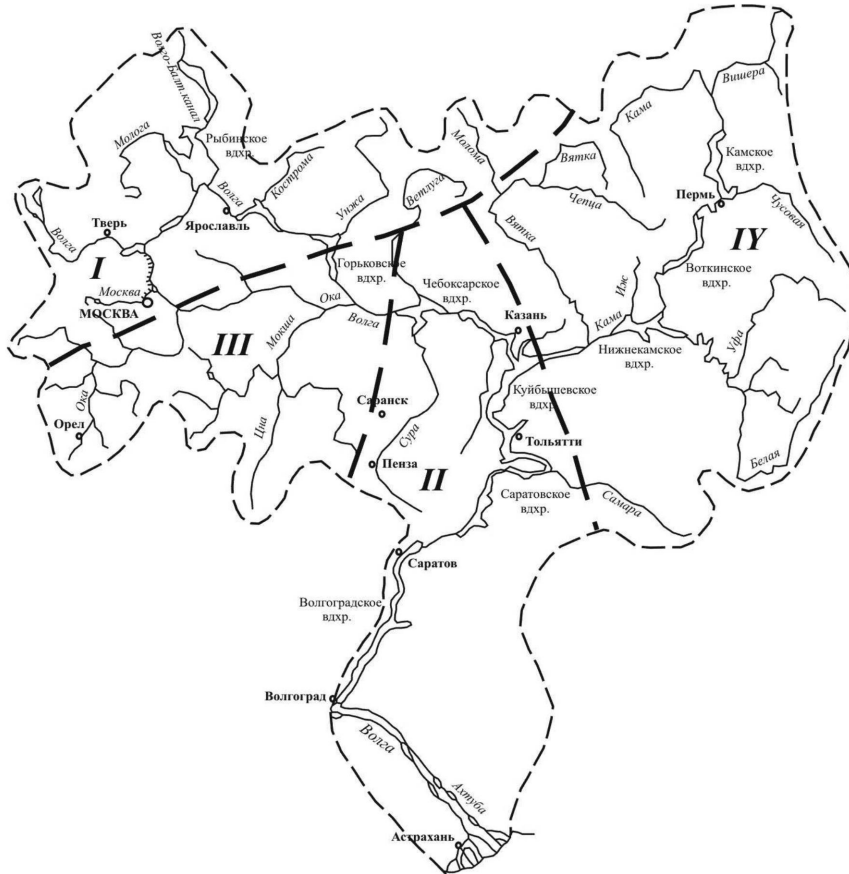


Рис. 1. Районирование бассейна Волги по характеристикам сезонной изменчивости речного стока

не несмещенных оценок корреляционных отношений (между группирующей и критериальными переменными). Для сравнения сходства классификаций вычисляется коэффициент Пирсона.

Сравнительный анализ показал, что выделенные по характеристикам сезонной изменчивости речного стока гидрологические районы на основе предложенных методов кластерного анализа не противоречат существующему районированию территории по типам водного режима рек и учитывают особенности формирования стока рек в разных физико-географических зонах. При этом статистические критерии оценки классификации для проведенного райони-

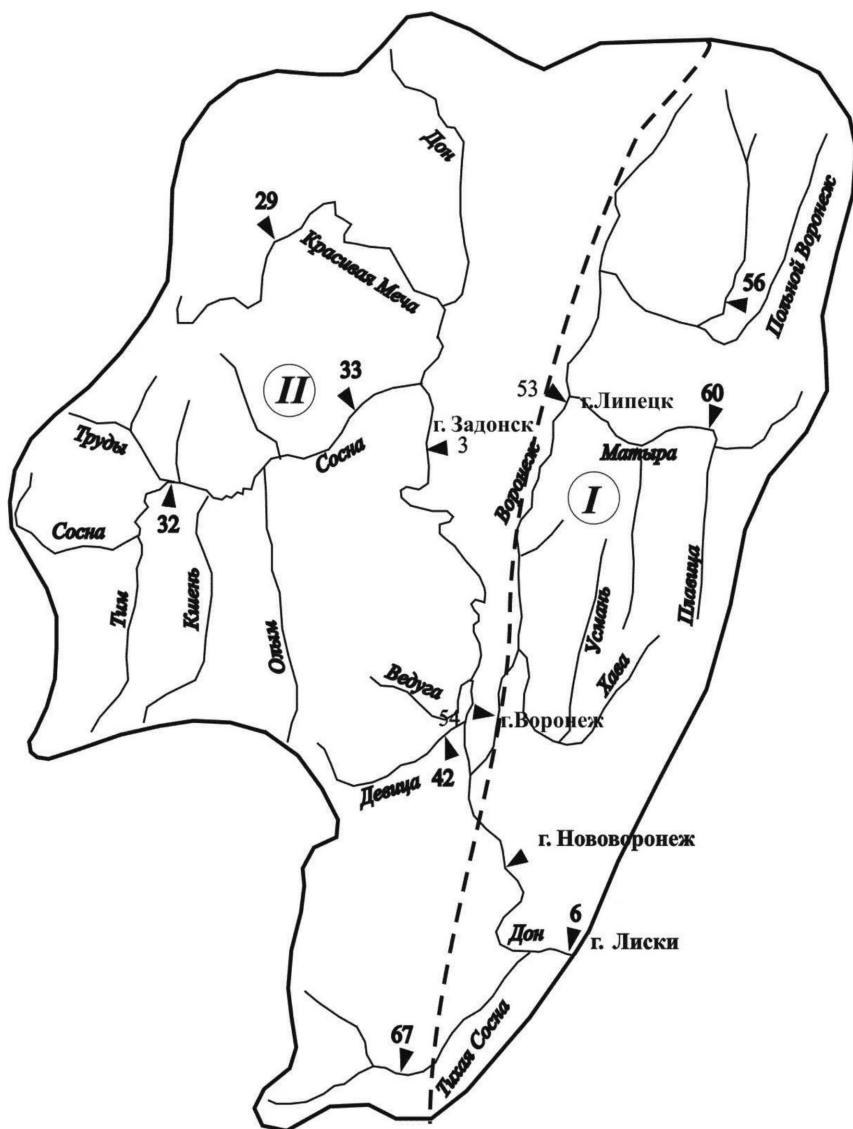


Рис. 2. Районирование бассейна верхнего Дона по характеристикам сезонной изменчивости речного стока

рования лучше, так как уменьшается количество групп, а также разброс (дисперсия) характеристик сезонного речного стока по объектам внутри них.

Классификация (группировка) рек позволила провести совместный анализ по группам однородных гидрологических объектов. Это дало возможность увеличить исходную информацию за счет ее соответствующего объединения и выявить возможные случайные ошибки. Обобщенные статистические параметры речного стока, полученные для таких групп, могут служить для его дальнейшего моделирования.

В связи с этим для каждого гидрологического района рассчитаны обобщенные (региональные) характеристики сезонных колебаний речного стока. Региональные оценки параметров стохастических моделей колебания речного стока получены на основе анализа всех репрезентативных рядов наблюдений на территории РФ. В результате для каждого однородного гидрологического района рассчитаны следующие обобщенные характеристики: параметры распределения речного стока по месяцам (коэффициенты вариации и их соотношение с коэффициентами асимметрии, автокорреляционные функции), спектральные функции распределения.

В частности, в пределах однородных гидрологических районов получены осредненные оценки параметров автокорреляционных функций с месячным интервалом дискретности для нормализованных последовательностей. Вследствие существенных различий в генезисе формирования стока для разных сезонов года корреляционная структура нормализованных последовательностей меняется для всех районов в зависимости от месяца года. Периоды устойчивой меженни характеризуются большими коэффициентами корреляции стока смежных месяцев и слабым затуханием автокорреляционной функции в целом, что естественным образом объясняется преобладанием в питании этого периода грунтовой (подземной) составляющей. Сток весенних (половодных) месяцев слабо связан с предшествующими расходами воды. Автокорреляционные функции половодных месяцев быстро затухают.

Выполненный совместный спектральный и корреляционный анализ временных рядов показал различную сложность стохастических моделей для рек с различным гидрологическим режимом. Выявлено, что количество значимых гармоник определяется сложностью гидрологического режима. Режим стока рек с преобладанием дожде-

вых паводков может быть воспроизведен с помощью более простой стохастической модели, чем режим рек с выраженным весенним половодьем. По-видимому, для полного описания асимметричного паводочного гидрографа требуется большее количество гармоник (в Фурье-анализе), что и отражается в форме спектральной плотности периодически коррелированного процесса.

Полученные обобщенные (региональные) показатели, отражающие пространственные закономерности сезонных колебаний речного стока, будут использованы в качестве параметров стохастических моделей речного стока.

Выводы

Разработанные стохастические модели и статистические методы оценивания позволили провести группирование гидрологических объектов наблюдений по сезонной изменчивости речного стока.

Применение оригинальных методов, связанных с анализом многомерных гидрологических данных, позволило получить новые варианты гидрологического районирования по характеру сезонных колебаний речного стока.

На основе многовариантных экспериментов с разными наборами показателей внутригодовых колебаний речного стока, полученных в процессе преобразования временных рядов среднемесячных расходов воды, проведено гидрологическое районирование ряда речных бассейнов на территории России. Регионализация производилась с помощью нового подхода к процедурам кластерного анализа с использованием регрессионной модели на основе радиальных базисных функций.

Даны оценки обобщенных (региональных) характеристик колебаний речного стока для каждого выделенного однородного гидрологического района. Полученные региональные показатели, отражающие пространственные закономерности сезонных колебаний речного стока на территории РФ, будут использованы в качестве параметров стохастических моделей речного стока.

Внедрение в гидрологические исследования фундаментальных результатов, полученных в области теории вероятностей и случайных процессов в последние десятилетия, способствует достоверности полученных результатов.

Литература

1. Болгов М.В., Мишон В.М., Сенцова Н.И. Современные проблемы оценки водных ресурсов и водообеспечения. М.: Наука, 2005. 318 с.
2. Енюков И.С., Ретинская И.В., Скуратов А.К. Статистический анализ и мониторинг научно-образовательных интернет-сетей. М.: Финансы и статистика, 2004. 320 с.
3. Кендалл М.Дж., Стюарт А. Многомерный статистический анализ и временные ряды. М.: Наука, 1976. 736 с.
4. Кузин П.С. Классификация рек и гидрологическое районирование СССР. Л.: Гидрометеиздат, 1960. 455 с.
5. Кузин П.С., Бабкин В.И. Географические закономерности гидрологического режима рек. Л.: Гидрометеиздат, 1979. 200 с.
6. Сенцова Н.И. Рациональное водопользование в бассейне Дона в условиях меняющихся антропогенных и климатических воздействий // Материалы Международной научно-практической конференции «Роль мелиорации и водного хозяйства в реализации национальных проектов». Часть 1. М.: МГУП, 2008. С. 402–408.
7. Сенцова Н.И. Пространственные закономерности сезонных колебаний речного стока в бассейне реки Волги: Сб. науч. трудов Всероссийской конференции «Водные проблемы крупных речных бассейнов и пути их решения». Барнаул, 2009. С. 204–216.

N.I. Sentsova

HYDROLOGICAL REGIONALIZATION OF THE RIVER BASINS ACCORDING TO SEASONAL VARIABILITY OF RIVER RUNOFF CHARACTERISTICS

Regionalization of the river basin according to characteristics of seasonal variability of a river runoff is considered. New series of river runoff characteristics are generated for classification on the basis of time series of monthly water discharges. As a result of multiple experiments sets of numerical parameters, which can be used for allocation of homogeneous hydrological areas, were obtained as a result of transformation of initial time series of monthly water discharges. These parameters include autocorrelation functions, indexes of seasonal prevalence and function of spectral density of monthly average water discharges.

Software system «TeleStat» destined for statistical analysis of multidimensional data and time processes was used for revealing of groups of homogeneous objects. Division into districts was made with the help of cluster analysis using regression model on the base of radial basic functions. Obtained sets of numerical parameters for each observational point were used as variables. Weights of variables were selected in order to provide the combination both geographical (coordinate), and criterion closeness of objects (hydrological gauges). Procedures of regionalization, which were carried out according to the character of seasonal fluctuations of river runoff, allowed to allocate homogeneous hydrological areas on the territory of some large Russia rivers, in particular Volga and the Upper Don.

Classification of the rivers gave the opportunity to carry out combined analysis on groups of homogeneous hydrological objects. Generalized statistical characteristics of river runoff fluctuations were calculated for each hydrological area. Such statistical parameters of river runoff can be used in its further modeling.