

6. Нефедова Т. Г. Сельская Россия на перепутье. М.: Новое изд-во, 2003. 408 с.
7. Регионы России: Стат. сб. В 2 т. Т. 2. М.: Госкомстат, 1998. 797 с.
8. Регионы России. Социально-экономические показатели - 2015 г. URL: // [http://www.gks.ru/bgd/regl/b15\\_14p/Main.htm](http://www.gks.ru/bgd/regl/b15_14p/Main.htm) (дата обращения: 07.12.2016).
9. Россия в цифрах, 2016 г. URL: // [http://www.gks.ru/bgd/regl/b16\\_11/Main.htm](http://www.gks.ru/bgd/regl/b16_11/Main.htm) (дата обращения: 07.12.2016).
10. Сельское хозяйство России: Стат. сб. /Госкомстат России. М., 1995. 503 с.
11. Сельское хозяйство, охота и охотничье хозяйство, лесоводство в России - 2015 г. URL: // [http://www.gks.ru/bgd/regl/b15\\_38/Main.htm](http://www.gks.ru/bgd/regl/b15_38/Main.htm) (дата обращения: 07.12.2016).

## **ОПЫТ МОРФОМЕТРИЧЕСКОЙ КЛАССИФИКАЦИИ РЕЛЬЕФА ТЕРРИТОРИИ БАСЕЙНА Р. СЕЙМ НА ОСНОВЕ НЕЙРОННЫХ СЕТЕЙ КОХОНЕНА<sup>18</sup>**

**С.В. Харченко<sup>1,2</sup>, С.С. Мухарамова<sup>1</sup>, О.П. Ермолаев<sup>1</sup>**

<sup>1-</sup> *Казанский (Приволжский) федеральный университет, Казань*

<sup>2-</sup> *Институт географии РАН, Москва*

e-mail: [har4enkkoff@rambler.ru](mailto:har4enkkoff@rambler.ru), [smukhara@gmail.com](mailto:smukhara@gmail.com), [oyermol@gmail.com](mailto:oyermol@gmail.com)

Аннотация. В работе делается попытка морфометрической классификации рельефа территории в бассейне р. Сейм. В качестве операционных единиц классификации используются речные бассейны 3 порядка по Стралеру-Философову и их межбассейновые пространства. Дается подробное поэтапное описание алгоритма. Используемый метод — построение нейронной сети Кохонена с последующей иерархической классификацией нейронов. Показан пример получаемых результатов. Выделены 5 классов, которые, за вычетом ошибочно выделенных (по причине несовершенства входных данных), можно интерпретировать как морфометрические районы на данной территории.

Ключевые слова: рельеф, речной бассейн, нейронные сети, самоорганизующиеся отображения Кохонена

Проблема автоматизированного картографирования рельефа, в первую очередь, его морфологии, но также и поиск на основе морфометрических параметров границ генетических комплексов — одна из перспективнейших и дискуссионных проблем геоморфологии. Несмотря на то, что, по мнению ряда исследователей, эта проблема окончательно не может быть решена в силу самой природы рельефа (в частности, геоморфологической конвергенции и гомологии), думается, сами попытки ее решения могут много дать теории геоморфологии. Такие попытки непрерывно предпринимаются исследовательскими группами по всему миру с большим или меньшим успехом [1,3,6,7,11 и др.]. Ими используется весь богатый методический

---

<sup>18</sup> Исследование выполнено за счет гранта Российского научного фонда (проект №15–17–10008)

аппарат математики (в основном, мат. статистики) и геоморфометрии. Разнообразие подходов, отсутствие четкого понимания того, как вообще подступаться к этой задаче, отвращает от нее многих специалистов. Однако не исключено, что в разнообразии подходов и методик следует видеть не проблему, а предпосылку к, наконец, успешному решению.

Одним из подходов является использование нейронных сетей. В картографии широкое распространение получили т.н. «самоорганизующиеся карты Кохонена» или «отображения Кохонена» по фамилии финского математика, впервые описавшего этот класс нейронных сетей [9]. Отдельный вопрос — выбор оптимальных операционных единиц классификации (ОЕК). ОЕК — это, своего рода, «атомы» проводящейся классификации, минимальные единицы территории, предполагаемые внутренне однородными. Распространено мнение, что для морфометрических классификаций удачным выбором являются водосборные бассейны низких рангов. Другое возможное решение — классификация по пикселям растров, что, однако, лишает возможности использовать такие статистики, как ранг, стандартное отклонение, минимум, максимум метрик рельефа. В этом случае используются только сами значения пикселей.

В данной работе делается попытка морфометрической классификации рельефа территории бассейна р. Сейм (Украина-Россия) по бассейнам более низкого ранга, примерно соответствующим третьему порядку по системе иерархии водотоков Стралера-Философова [4,12], и их межбассейновым пространствам. Охватываемая в исследовании площадь территории (бассейн р. Сейм) составляет 27 900 км<sup>2</sup>. На этой площади расположено 408 бассейнов третьего порядка и их межбассейновых пространств со средней площадью чуть более 68 км<sup>2</sup>. Именно эти 408 участков использованы в качестве ОЕК.

Для классификации полезными были сочтены следующие метрики рельефа (не все из них использовались в дальнейшем): абсолютная высота, крутизна, общая (тангенциальная) кривизна, «южность» (косинус экспозиции поверхности, взятый со знаком минус), показатель суммарного стока (площадь водосбора для каждой ячейки растра), TPI (topographic position index [8,13]), TRI (terrain roughness index [10]). Все метрики рассчитаны по сетке 50 м. Исходные данные для расчета — SRTM 1". Следует отметить, что подбор показательных метрик — это, пожалуй, важнейший этап для проведения успешной классификации. И широкое обсуждение репрезентативного набора метрик — уже насущная потребность для специалистов, занимающихся подобными задачами. Для каждой из перечисленных метрик рассчитывались следующие статистики: среднее, размах, минимум, максимум, стандартное отклонение. Все расчеты произведены средствами языка R с привлечением дополнительных пакетов «kohonen», «maptools», «raster», а также пакета «ggplot2» для визуализации результатов.

Использованный алгоритм морфометрической классификации включает в себя следующие этапы:

1) подготовка набора растров на исследуемую территорию, отражающих изменчивость тех морфометрических характеристик рельефа, которые предполагается использовать для классификации (например, высота, крутизна, «южность» и т.д.);

2) подготовка векторного слоя ОЕК (например, речных бассейнов, клеток квадратной сетки, административно-территориальных образований);

3) проверка распределений морфометрических характеристик на близость к нормальному; в случае необходимости — трансформация растров (логарифмическая, степенная, Тьюки) для приведения распределения, по возможности, к около-нормальному;

4) извлечение зональной статистики морфометрических характеристик для «зон» операционных единиц классификации. При этом для разных морфометрических характеристик рассчитываться могут различные статистики. Например, для показателя абсолютной высоты есть смысл рассчитывать среднее значение и размах значений в контурах ОЕК, для крутизны — средний и максимальных уклон и т.д.;

5) «масштабирование» извлеченных характеристик. Т.к. разные метрики рельефа имеют часто несопоставимые характерные значения (допустим, высокое плато с характерными уклонами  $0-8^\circ$  и отметками  $> 1000$  м), зональные статистики необходимо привести к виду, где среднее значение будет равняться 0, а стандартное отклонение — 1. В математической статистике эта процедура носит название Z-преобразования, Z-шкалирования или стандартизации;

6) «взвешивание» характеристик. Если исследователь считает, что входные характеристики должны иметь разные роли в процедуре классификации, требуется умножить уже масштабированные ряды значений по каждой метрике на «вес» этой метрики. По умолчанию, «веса» характеристик принимаются равными единице;

7) обучение модели. Создается шаблон для расчета параметров отображения Кохонена в виде сетки с квадратными или гексагональными ячейками с заданными размерами (количеством ячеек). Затем с этим шаблоном итеративно сравниваются все объекты (ОЕК), для каждого из которых находится наиболее похожий на него нейрон (ячейка). После нахождения наиболее похожего нейрона, его положение в сети смещается в сторону точки с координатами, соответствующими значениям (метрикам) данного ОЕК. Таким образом, сеть деформируется, подстраиваясь по реальную структуру данных; близкородственные нейроны группируются рядом друг с другом, и наоборот — ОЕК с сильно отличающимися характеристиками оттягивают соответствующие им нейроны в противоположные стороны.

8) когда модель обучена (и если это сделано на репрезентативной выборке), при добавлении новых объектов можно сразу относить их к тому или иному нейрону. Обычно, для получения более детальной картины используются сети с достаточно большим, намеренно избыточным

количеством нейронов — например, 15\*15 или 10\*10 ячеек и т.п. Это приводит к тому, что схожие объекты оказываются в разных, хотя и близко расположенных нейронах. Для упрощения работы можно отказаться от такого количества нейронов и классифицировать их, объединив в относительно небольшое кол-во классов / кластеров (например, до 10). Обычно для этого используется иерархическая классификация по методу Уорда или другие аналогичные подходы. Уточним, что здесь и далее термины «класс» и «кластер» понимаются нами как синонимы, хотя на этот счет имеются разные мнения.

9) в результате для каждой ОЕК получаем номер нейрона, с характерными для него параметрами метрик рельефа, а также номер класса более высокого уровня обобщения, к которому отнесен нейрон. По этим данным формируется карта/схема, где методом качественного фона показаны различные ОЕК, отличающиеся номером класса/нейрона, и, следовательно, при удачно подобранных входных метриках — морфологией, а, возможно, и генезисом рельефа (рис.1).

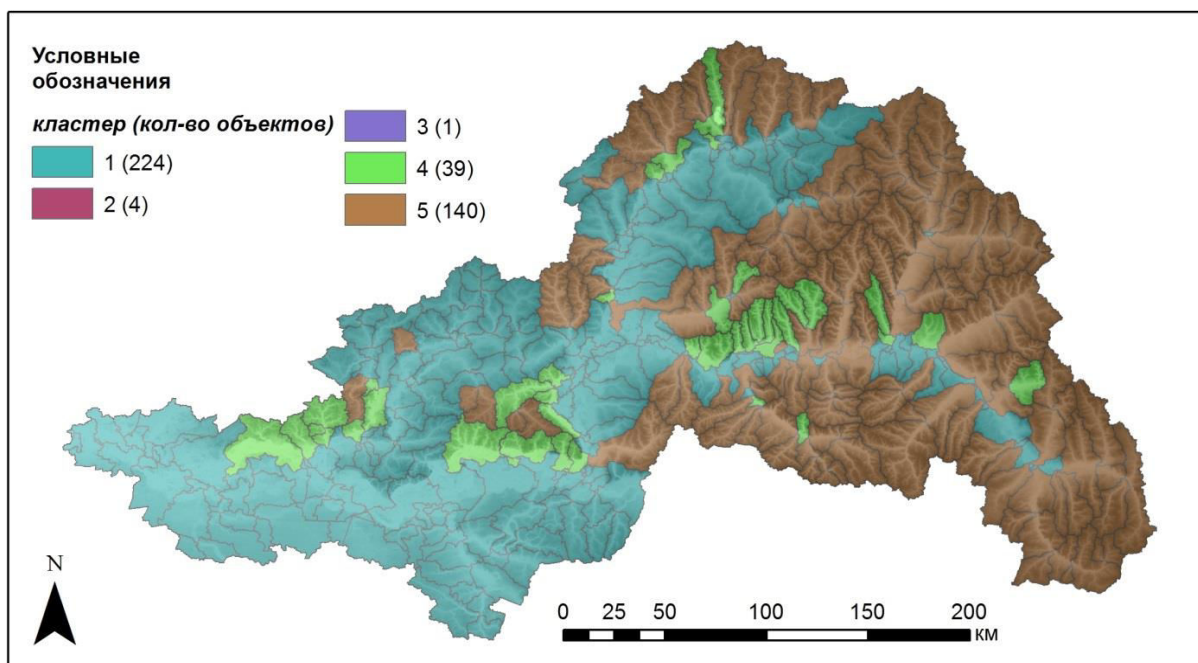


Рис.1. Пример морфометрической классификации рельефа бассейна р. Сейм

В данном примере нами было выделено 5 кластеров. Входными параметрами для классификации были взяты: средняя высота, размах высот и их стандартное отклонение, средняя крутизна, средний суммарный сток, стандартное отклонение общей кривизны и средние значения ТРІ и ТRІ. Количество объектов в выделенных классах сильно отличается. «Необходимость» выделения немногочисленных кластеров обусловлена их сильным отличием от соседних участков. Уменьшая количество кластеров с целью избавиться от немногочисленных и не занимающих значимых площадей кластеров №№ 2 и 3 за счет объединения близкородственных

кластеров, можно попасть в ситуацию, когда объединятся, например, кластеры №№ 4 и 5, а не нужные нам.

Сама структура исходных данных приводит к тому, что возникают такие редкие, но «яркие» по своей специфике кластеры, и устранять их либо же вовсе нет смысла, либо же стоит это делать за счет подбора более корректного пула исходных метрик. Например, кластер №2 — участки небольших межбассейновых пространств, возникающие в случае последовательного впадения притоков в главную реку на небольшом ее отрезке. Как результат, эти участки целиком или почти целиком умещаются в днище долин р. Тускарь, р. Псел и имеют относительно низкие показатели средней высоты, ее размаха и стандартного отклонения, весьма низкие средние уклоны и другие характерные черты, отличающие данный кластер от остальных больше, чем, например, кластеры №№ 4 и 5 отличаются между собой.

Самостоятельной задачей является качественная интерпретация выделенных кластеров, которую можно осуществлять при помощи подложенной цифровой модели рельефа (как на рисунке). Так, к кластеру №1 были отнесены, очевидно, низкие днища долин в верховьях крупнейших рек (рр. Сейм и Свапа), а также пониженные слабо расчлененные междуречья в среднем и нижнем их течении. Кластер № 2 уже был описан нами выше. Кластер №3, целиком состоящий только из одного объекта (межбассейновое пространство в районе г. Курчатов), выделяется аномальными показателями ТРІ, что, в свою очередь, обусловлено наличием крупных сооружений Курской АЭС, которые на цифровой модели рельефа отражаются в виде высоких, крутосклонных холмов на фоне пониженной поверхности днища речной долины. В зависимости от конкретной задачи исследования, такие «антропогенно-обусловленные» кластеры следует либо исключать из анализа, либо, напротив, уделять им повышенное внимание. Кластер №4 — приподнятые, сильно расчлененные бассейны с высокими средними показателями ТРІ и ТRІ, зачастую захватывающие значительные участки днищ долин. Кластер №5 — приподнятые, умеренно расчлененные бассейны с меньшими, нежели в предыдущем кластере, средними показателями ТРІ и ТRІ.

Более детальный анализ позволяет делать рассмотрение соответствующих диаграмм (рис.2). Пять классов, отраженных на рис.1, показаны и здесь теми же цветами.

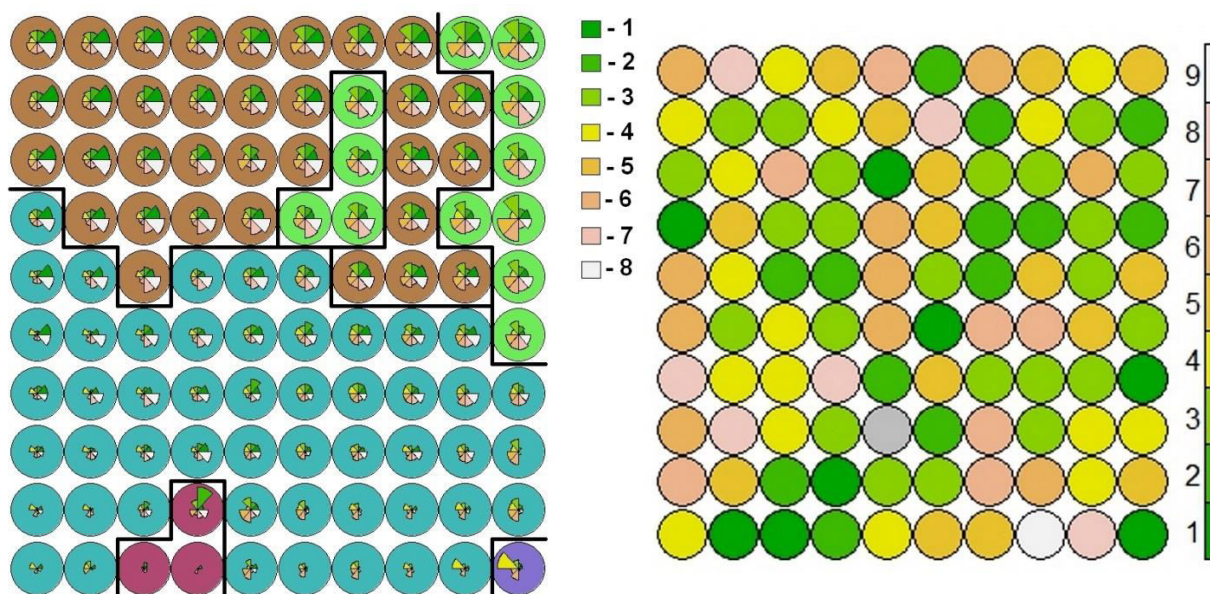


Рис.2. Слева — классифицированная нейронная сеть Кохонена. Для каждого класса показаны характерные относительные значения по 8 метрикам (1–5 — средняя высота, суммарный сток, уклон, ТРІ и ТRІ; 6–7 — стандартное отклонение кривизны и высоты; 8 — размах высот). Справа — количество объектов, отнесенных к каждому из нейронов.

На левой диаграмме для каждого нейрона показаны относительные величины параметров рельефа. Так, например, хорошо видно, что средний ТРІ в классе № 3 для единственного нейрона (см. слева) и единственного объекта в этом нейроне (см. справа) гораздо выше, чем по всем остальным нейронам. Уже, по-видимому, одно это обусловило выделение для него самостоятельного класса. Даже по самому размеру роз-диаграмм хорошо заметны отличия кластеров по относительным значениям метрик. Так, кластер №1 характеризуется относительно небольшими «розами» и, следовательно, невысокими значениями всех показателей. В противоположность ему, большие «розы» характерны для кластеров №№ 4 и 5.

Совершенно очевидно, что этот вариант классификации — лишь один из многих возможных и вряд ли самый удачный из них. Для повышения эффективности этого метода требуется решение еще нескольких сложных задач. Во-первых, определить набор репрезентативных морфометрических характеристик, способных отразить различия в морфологии и генезисе рельефа. Скорее всего, это будут не только классические общераспространенные метрики, но и сложные метрики, например, гармонические (характеризующие периодичность в структуре ЗП) [5]. Возможно, что-то нужно будет заимствовать из теории распознавания образов [2]. В конце концов, если специалист может увидеть на гипсометрической карте друмлины или конечно-моренные гряды, то потенциально может «увидеть» их и компьютер. Во-вторых, нужно определиться с оптимальной операционной единицей классификации — должны ли быть это речные бассейны и, если «да», то какого порядка? Если «нет», то что другое использовать в качестве ОЕК? В-третьих, стоит задача

ориентированного поиска ответов на поднятые выше вопросы. Если решается проблема выделения по различиям морфометрических характеристик генетических типов рельефа, то можно попытаться автоматизировать поиск оптимального решения — делать множество классификаций с разными входными метриками, их статистиками и «весами», разным количеством кластеров, а затем полученные границы по каким-либо формальным критериям сравнивать с желаемым результатом, например, с контурами векторной геоморфологической карты. Набор входных данных и параметров, который позволит получить наилучшее сходство, следует считать наиболее удачным и репрезентативным.

## Литература

1. Ермолаев О.П., Селиванов Р.Н. Автоматизированное ландшафтно-экологическое картографирование городских территорий с использованием нейронных сетей // Ученые записки Казанского университета. Естественные науки. — 2010. — Том 152, кн. 4. — С. 52–67.
  2. Иванов В.В., Чалова Е.Р. Опыт систематизации картографических образов геоморфологических объектов // Геоморфология. — 1987. — № 2. — С. 62–66.
  3. Мальцев К.А. Морфометрический анализ рельефа Республики Татарстан средствами ГИС-технологий. Автореферат ... канд. геогр. наук. — Казань: КПФУ, 2006. — 23 с.
  4. Философов В.П. Краткое руководство по морфометрическому методу поисков тектонических структур. — Саратов: Изд-во Саратов. ун-та, 1960. — 68 с.
  5. Харченко С.В. К вопросу о применении гармонического анализа при количественной характеристике рельефа // Геоморфология. — 2017. — № 2. — С. 14–24.
  6. Arrell K.E., Fisher P.F., Tate N.J., Bastin L. A fuzzy c-means classification of elevation derivatives to extract the morphometric classification of landforms in Snowdonia, Wales // Computers & Geosciences. — 2007. — No. 33. — Pp. 1366–1381.
  7. Dikau R., Brabb E.E., Mark R.K., Pike R.J. Morphometric Landform Analysis of New Mexico // Zeitschrift für Geomorphologie, Supplementband. — 1995. — No. 101. — Pp. 109–126.
  8. Gallant J.C., Wilson J.P. Primary topographic attributes // Terrain Analysis: Principles and Applications. — New York: Wiley, 2000. — Pp. 51–85.
  9. Kohonen T. Self-Organizing Maps. — Heidelberg: Springer, 1997. — 260 p.
  10. Riley S.J., De Gloria S.D., Elliot R. A Terrain Ruggedness that Quantifies Topographic Heterogeneity // Intermountain Journal of Science. — 1999. — Vol. 5, No.1-4. — Pp. 23–27.
  11. Seif A. Landform Classification by Slope Position classes // Bull. Env. Pharmacol. Life Sci. — 2014. — Vol 3 [11]. — Pp. 62–69.
  12. Strahler A.N. Quantitative Analysis of Watershed Geomorphology. — Transactions. American Geophysical Union. — 1957. — V. 38, №6 — Pp. 913-920.
- Weiss A.D. Topographic position and landforms analysis. Poster Presentation, ESRI Users Conference, San Diego, CA, 2001.