

Тестовые полигоны для диагностики состояния геосистем и развития методов интерпретации данных дистанционного зондирования Земли

А. А. Ямашкин , С. А. Ямашкин, М. Р. Байчурин, А. А. Токарев, И. С. Лямзина

*Национальный исследовательский Мордовский государственный университет,
Российская Федерация
(430005, г. Саранск, ул. Большевикская, 68)*

Аннотация. Исследование, представленное в статье, направлено на разработку методов и алгоритмов системного анализа данных тестовых полигонов с целью диагностики состояния геосистем и развития методов интерпретации данных дистанционного зондирования Земли.

Материалы и методы. Разработка методов и алгоритмов анализа и интеграции пространственной информации опиралась на анализ системы тестовых полигонов, раскрывающих особенности взаимодействия лесостепных и лесных геосистем Приволжской возвышенности и Окско-Донской низменности. Общая схема процесса составления цифровой карты геосистем реализуется посредством решения задач сбора, систематизации и анализа пространственных данных с построением иерархии геосистем с оценкой результатов моделирования для достижения цели практического использования пространственной информации.

Результаты и обсуждение. Объединение моделей в ансамбль на основе предложенной архитектуры метаклассификатора позволяет повысить устойчивость анализирующей системы: точность решений, принимаемых ансамблем, имеет тенденцию стремиться к точности наиболее эффективного моноклассификатора системы. Системный анализ дескрипторов территории, интегрируемых на основе данных из разных источников, дает существенный прирост точности классификации метагеосистем. Важно то, что картограммы представленных дескрипторов хорошо интерпретируются специалистами в области анализа данных в науках о Земле.

Заключение. Использование ансамблей, выстраиваемых по предложенной в статье методике, позволяет проводить оперативный автоматизированный анализ пространственных данных для решения задачи тематического картографирования метагеосистем и природных процессов. Расчет и консолидация территориальных дескрипторов позволяют снизить размерность анализируемых данных, облегчить допустимую емкость модели машинного обучения, повысить ее устойчивость к переобучению, не допустить значительного снижения точности классификации в рамках конкретной решаемой задачи.

Ключевые слова: метагеосистемы, пространственные данные, тестовые полигоны, территориальные дескрипторы, ансамбли.

Источник финансирования: Исследование выполнено за счет гранта Российского научного фонда № 22-27-00651, <https://rscf.ru/project/22-27-00651/>.


Для цитирования: Ямашкин А. А., Ямашкин С. А., Байчурин М. Р., Токарев А. А., Лямзина И. С. Тестовые полигоны для диагностики состояния геосистем и развития методов интерпретации данных дистанционного зондирования Земли // *Вестник Воронежского государственного университета. Серия: География. Геоэкология*, 2022, № 4, с. 4-18. DOI: <https://doi.org/10.17308/geo/1609-0683/2022/4/4-18>

ВВЕДЕНИЕ

Важнейшими направлениями современных исследований наук о Земле являются разработка методологии и систем методов диагностики состояния геосистем, прогнозирование развития и мини-

мизации (предупреждения) проявления природных и природно-техногенных чрезвычайных экологических ситуаций. Решение этой задачи предполагает формирование системы тестовых полигонов, которые оптимизируют поиск закономерностей

© Ямашкин А. А., Ямашкин С. А., Байчурин М. Р., Токарев А. А., Лямзина И. С., 2022

 Ямашкин Анатолий Александрович, e-mail: yamashkin56@mail.ru

развития геоэкологических процессов на глобальном, региональном и локальном уровнях.

Исследование, представленное в статье, направлено на разработку методов и алгоритмов системного анализа данных тестовых полигонов с целью диагностики состояния геосистем и развития методов интерпретации данных дистанционного зондирования Земли. В качестве основных объектов исследования выступают геосистемы – «...особый класс управляющих систем; земное пространство всех размерностей, где отдельные компоненты природы находятся в системной связи друг с другом и как определенная целостность взаимодействуют с космической сферой и человеческим обществом» [2]. Важнейшими свойствами геосистем при этом являются территориальность, структурность, иерархичность, взаимосвязанность системы и среды, качественная и количественная сложность, эмерджентность, динамичность, устойчивость, целостность, автономность, управляемость. В совокупности они определяют метакронный характер проявления геоэкологических процессов на региональном и локальном уровнях.

Множественность описаний геосистем показывает, что в качестве основного инструментария исследования состояний геосистем и их изменения должны выступать ГИС-технологии, создающие информационное пространство, оптимизирующее организацию геоэкологического мониторинга состояния многокомпонентных образований географической оболочки. Они являются надежным инструментом для решения широкого спектра научно-практических задач анализа, оценки, прогнозирования и интеллектуальной поддержки управленческих решений в сфере мониторинга состояния геосистем.

МАТЕРИАЛЫ И МЕТОДЫ

В организации исследований геосистем целесообразно выделить следующие этапы: обоснование проблемных ситуаций, подготовка исходных данных и разработка алгоритмов их слияния, типологическая систематизация информации и составление синтетической карты геосистем, ансамблевый анализ восприимчивости геосистем к внешним воздействиям, генерация выходных данных и использование полученных результатов. Общий алгоритм составления электронных карт геосистем тестовых полигонов представлен на рисунке 1.

Обоснование проблемных ситуаций. Важнейшими этапами создания синтетической карты геосистем являются выделение проблемных ситуаций, связанных с развитием геоэкологических

процессов, создание системы тестовых полигонов регионального и локального уровней для диагностики структуры, функционирования, динамики и развития геосистем, формирование исходных данных, моделирование пространственно-временной структуры геосистем, генерация выходных данных, анализ результатов и их использование для принятия управленческих решений.

Так, в схеме физико-географического районирования Республики Мордовия система тестовых полигонов ориентирована на раскрытие пространственно-временной организации геосистем в зоне взаимодействия лесостепи Приволжской возвышенности и смешанных лесов Окско-Донской низменности. Хозяйственное освоение региона сопряжено с актуальностью принятия управленческих решений в области регулирования водного баланса и оптимизации проблем водоснабжения, минимизации развития экзогеодинамических процессов, сохранения почвенного плодородия и биологического разнообразия.

Подготовка данных. Спектр проблемных ситуаций определяет структуру баз данных, отражающих структуру геосистем на региональном и локальном уровнях. Региональная ГИС «Мордовия» включает систему электронных карт с информацией по основным анализируемым элементам геосистем, данные дистанционного зондирования Земли (ДЗЗ).

Применение тестовых полигонов актуально с позиции решения задачи консолидации информативного набора обучающих выборок данных для работы автоматизированных алгоритмов, точность которых можно значительно увеличить, откалибровав процесс классификации на основе методом наземных измерений в ходе полевых исследований, предполагающих непосредственное наблюдение объектов. В рамках тестовых полигонов рассматриваются морфологические части природных и антропогенных ландшафтов в границах отдельных типов земель. Тестовые полигоны должны удовлетворять набору качественных требований: 1) достаточное структурное разнообразие, предполагающее наличие в рамках исследуемой территории ландшафтов, достаточных для создания исчерпывающих наборов эталонных выборок, характеризующих формы рельефа, почвы, водные объекты, растительный покров, антропогенные объекты; 2) высокое качество регистрируемых пространственно ассоциированных характеристик, обеспечивающих формирование достаточного набора стабильных излу-

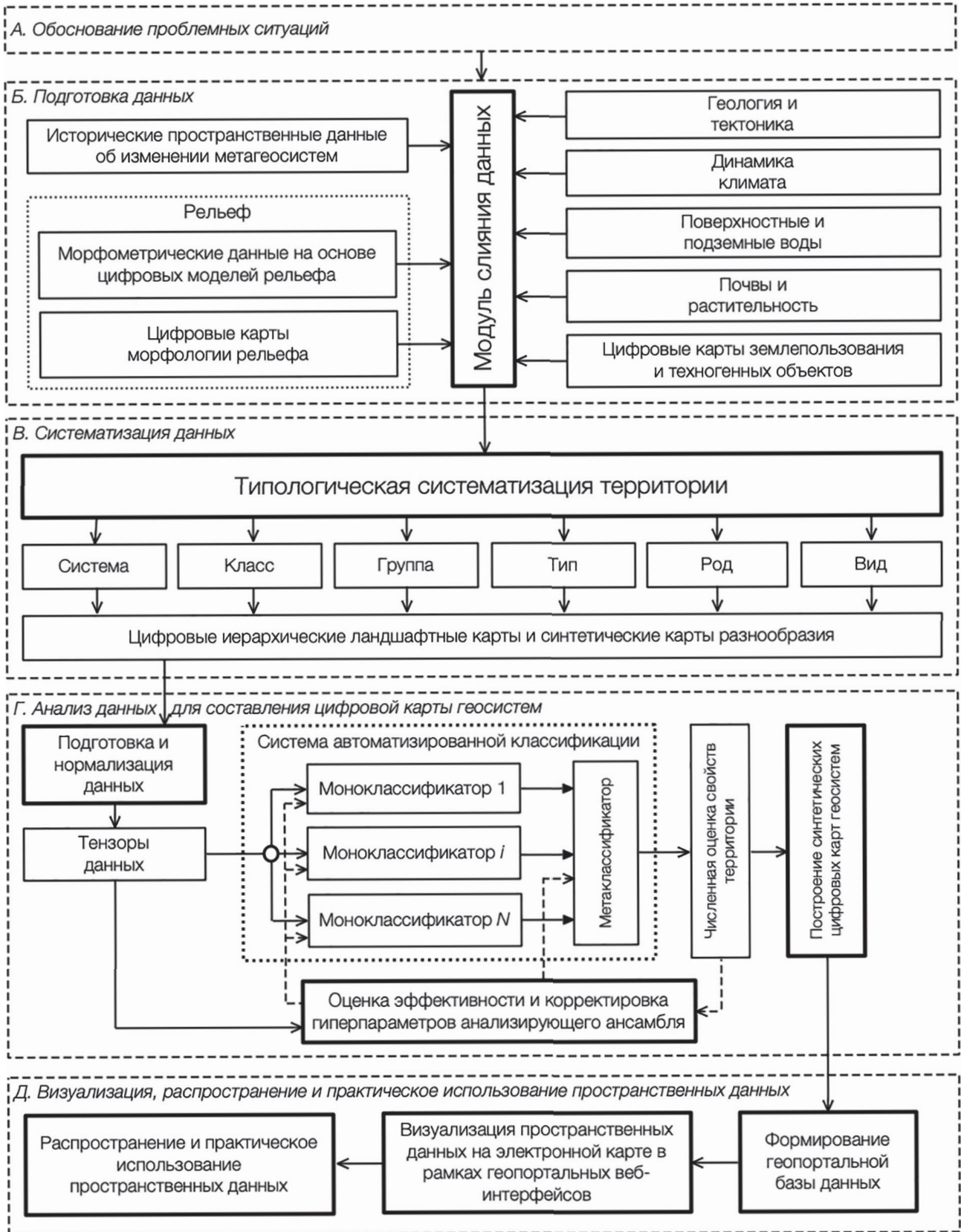


Рис. 1. Алгоритм составления цифровой карты геосистем [Fig. 1. Algorithm for compiling a digital map of geosystems]

тельно-отражательных характеристик геосистем; 3) достаточные для репрезентативной выборки пространственных данных расположение и размеры полигона, определяющие статистически достаточное число представленных классов объектов; 4) корректность методического, нормативного и технологического обеспечения процесса консолидации пространственных данных, содержащих объективные и точные сведения о пространственно-временной организации геосистем; 5) информативность метаописания, представляющего важные сведения о пространственных данных, включая информацию о съемке, описании метагеосистем, административную характеристику.

Тестовые полигоны, удовлетворяющие представленным качественным требованиям, могут быть использованы для оптимизации моделей автоматизированной классификации, основанных на применении технологий машинного обучения и могут быть объединены в единую систему для повышения качества проектных работ в области картографирования ландшафтов, управления хозяйством, ведения точного земледелия, разведки месторождений природных ресурсов, мониторинга природных и природно-техногенных стихийных процессов.

Систематизация данных. В качестве центрального звена региональной географической информационной системы выступает синтетическая карта геосистем, опирающаяся на широкое использование многозональных космических снимков. Она является базовой основой для: 1) составления (ревизии) тематических карт, раскрывающих структуру геосистем и развитие геоэкологических процессов; 2) выявления зон влияния геотехнических систем на состояние окружающей среды; 3) оценки негативных последствий прямого и косвенного техногенного воздействия на геосистемы; 4) прогнозирования динамики изменения геоэкологической обстановки; 5) принятия управленческих решений по предотвращению (минимизации) развития деструктивных геоэкологических процессов.

Пространственно-временную структуру геосистем можно представить в виде совокупности процессов перемещения, обмена и трансформации энергии, вещества и информации между ее элементами и окружающим географическим пространством. Определяющими факторами формирования структуры, развития, динамики и функционирования геосистем являются: макроклиматические фак-

торы, тектонические формы рельефа, особенности водного и геохимического режима, почвенно-биологические и экзогеодинамические процессы, растительные сообщества. В соответствии с этим в работе используется типологическая классификация геосистем по В. А. Николаеву [1]. При синтетическом картографировании геосистем предусматривается следующая иерархия таксономических единиц: разряд, подразряд, класс, подкласс, тип, подтип, род, подрод ландшафтов.

Использование пространственных данных разной степени генерализации способствует установлению закономерностей пространственно-временной организации геосистем, увеличивает надежность дешифрирования, способствует более точной интерпретации диагностических признаков.

Стадия *«Анализ данных для составления цифровой карты геосистем»* направлена на осуществление автоматизированного анализа данных пространственного мониторинга с применением вычислительных алгоритмов [4]. Программные и математические модели, группируемые в ансамбли, функционируют на основе выделения статистических закономерностей и эталонных признаков пространственно распределенных объектов [6].

Методы и алгоритмы машинного обучения могут быть эффективно использованы для интерпретации геопространственных данных, для которых характерны свойства пространственной зависимости, пространственной неоднородности и масштабируемости. Одновременно с этим применение методов и алгоритмов глубокого машинного обучения к анализу геопространственных данных сталкивается с рядом открытых проблем, требующих научно обоснованного решения. Модели машинного обучения, используемые для решения задачи классификации геосистем, могут иметь различную архитектуру (искусственные нейронные сети, деревья принятия решений, машины опорных векторов) и гиперпараметры. Более того, они способны успешно обучаться на разных наборах данных об интерпретируемой территории, которые могут быть многомерными и мультимодельными.

В качестве объекта системного пространственного анализа в современной науке в области анализа пространственных данных выступают геосистемы [2]. Учение о геосистемах актуально как в сфере исследования природных объектов и процессов, так и в области анализа их взаимодействия с социальными и экономическими системами. В этом случае необходимо ввести

понятие «метагеосистемы», цифровые модели территории как основного объекта пространственного анализа [3].

В качестве завершающей ступени выделяется этап «**Визуализация, распространение и практическое использования полученных результатов**». Для его реализации необходимо обеспечить формирование мультимодельной геопортальной пространственной базы данных. Решение задачи визуализации пространственных данных достигается посредством цифровых карт, функционирующих в рамках геопортальных веб-интерфейсов, и позволяющих достичь цели распространения и практического использования данных о региональных метагеосистемах.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Пространственно-временная структура тестовых полигонов «Мордовия» определяется географическим положением в системе суббореальных семиаридных геосистем пластово-ярусной Приволжской возвышенности и пластовой Окско-Донской низменности, что выражается в функционировании лесостепных ландшафтов и генетически, территориально сопряженных с ними лесных, луговых, болотных и других.

Разряд (подразряд). Макроклиматические факторы: радиационный режим, характер атмосферной циркуляции, водный и тепловой баланс, соотношение тепла и влаги, определяющие характер развития физико-географических процессов в геосистемах. Макроклиматический фактор обуславливает формирование широтной ландшафтной поясности и долготной ландшафтной секторности, определяющих общие черты хозяйственного освоения геосистем и зональные особенности землеустройства.

Радиационный баланс составляет за год 1638,29 МДж/м², в декабре он равен –16,76 МДж/м², в июне – +339,39 МДж/м². Средняя температура воздуха за год равна 3,5 – 4,0 °С, в годовом ходе она меняется от 19,0 – 19,8 °С в июле до –11,2 –11,7 °С в январе. Относительная влажность воздуха изменяется от 84–86 % в декабре–январе до 62–64 % в мае–июне и в среднем за год составляет 76–77 %. Осадков в среднем за год выпадает 598–636 мм, из них 221–252 мм – в холодный период, с ноября по март, и 337–391 мм – в теплый период, с апреля по октябрь. В среднем в году наблюдается 70 дней с жидкими осадками (342 мм).

Класс (подкласс). Морфотектонические структуры, отличающиеся по возрасту, свойствам литологического состава коренных горных пород и

четвертичных отложений, формируют литогенную основу природных комплексов, литогидрогенные системы, определяют развитие экзогеодинамических процессов, ведущих к проявлениям высотной смены геосистем. В классе равнинных геосистем исследуемой территории выделяются следующие подклассы: 1) возвышенная эрозионно-денудационная равнина с фрагментами олигоценовой поверхности выравнивания на абсолютных отметках 280–320 м с выходами на дневную поверхность кремнисто-карбонатных пород палеоцена (опоками, песчаниками, мергелями, песками) и карбонатных пород верхнемелового возраста (опоки, мел); глубина эрозионного вреза до 100–120 м; 2) эрозионно-аккумулятивная равнина с фрагментами плиоцен-плейстоценовых поверхностей выравнивания на приводораздельных пространствах с абсолютными отметками 230–250 м, сложенная терригенными мезозойскими породами, перекрываемыми моренными и элювиальными суглинками; 3) низменные водно-ледниковые пологоволнистые равнины, сложенные гляциальными образованиями, с абсолютными отметками 180–210 м, характеризующиеся широкими водоразделами – до 8–10 км, пологими и слаборасчлененными склонами; глубина эрозионного вреза не превышает 30–40 м; 4) низинные геосистемы долин рек зон линейных тектонических структур; структурные линии часто являются зонами разгрузки подземных вод, повышенной активности экзогеодинамических процессов, связанных с геологической средой: карстовых, оползневых, суффозионных и других.

Группы (подгруппы) геосистем, выделяемые по водному режиму, формирующиеся в результате сложных сочетаний климатических и литолого-геоморфологических факторов, определяют развитие водно-геохимических процессов, обеспечивающих обмен веществом и энергией между ярусами геохимического ландшафта, формирование пространственных закономерностей изменения трофности и гидроморфности геосистем. Основной источник водных ресурсов – атмосферные осадки, которые не только образуют речной и поверхностный сток, валовое увлажнение почв, испарение, но и обеспечивают запасы подземных вод. Общая структура литогидрогенных систем индицируется по зонам разгрузки подземных вод.

Особое значение в функционировании литогидрогенных геосистем древних ложбин стока имеет водоносный каменноугольно-пермский карбонатный комплекс, основные области пита-

ния которого на территории Республики Мордовия приурочены к областям Алатырского, Окско-Цнинского и Сурско-Мокшинского валов, где в сводовых частях известняки и доломиты верхнего карбона и нижней перми выходят на поверхность, – в бассейнах рек Сатис, Уркат, в среднем течении Мокши, у села Сивинь в долине Сивини, в центральной части бассейна реки Вад. Мощность обводненной толщи с пресными водами составляет от 20 до 250 м. Коэффициент фильтрации изменя-

ется в значительных пределах – от 3,3 до 80 м/сут. Наиболее высокими фильтрационными свойствами модельный слой обладает в области структурных поднятий. С глубиной трещиноватость затухает и водообильность толщи падает.

Типы (подтипы) геосистем выделяются по почвенно-биологическим признакам (рис. 2). Этот уровень ландшафтной дифференциации выражается в проявлении зональности, интразональности, экстразональности.

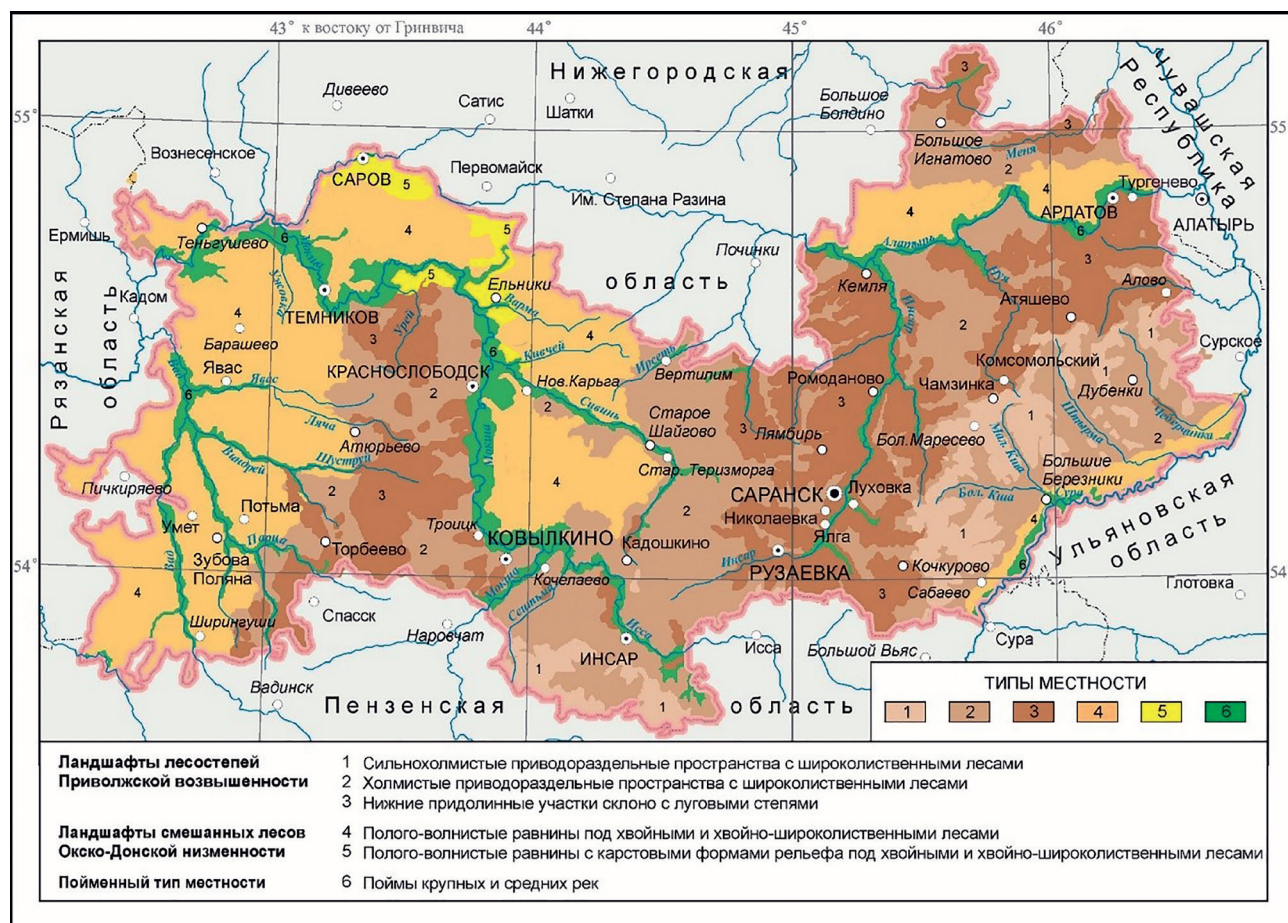


Рис. 2. Типы геосистем Республики Мордовия
[Fig. 2. Types of geosystems of the Republic of Mordovia]

Особенности взаимодействия зональных и азональных факторов определяют проявление на тестовом полигоне высотной мезозональности геосистем: хвойных и смешанных лесов с дерново-подзолистыми и серыми лесными почвами супесчаного и легкосуглинистого механического состава, с очаговым освоением; луговых степей с доминированием в структуре почвенного покрова выщелоченных черноземов, значительно освоенных; широколиственных лесов на темно-серых и серых лесных почвах, выборочно

освоенных; дубрав на серых лесных щебнистых почвах, выборочно освоенных.

В лугово-степных геосистемах характерны значительные массивы пашенных ландшафтов с редкой сетью крупных земледельческих поселений, в типах геосистем смешанных лесов общей чертой культурного ландшафта является очаговое земледельческое и лесохозяйственное освоение с немногочисленными сельскими и лесными поселениями.

Роды (подроды) геосистем определяются развитием экзогеодинамических процессов (эрозия,

карст, суффозия и т.п.), определивших морфоскульптурные формы рельефа и распределение генетических типов четвертичных отложений. Различия в литогенной основе определяют формирование микро- и мезоклиматических условий, особенности водного и геохимического режимов, структуру почвенных комбинаций, а, следовательно, развитие процессов хозяйственного освоения.

Виды геосистем отражают структуру местопроизрастаний и сообщества растений. Анализ иерархической структуры геосистем

показывает, что для разработки методов и алгоритмов анализа и интеграции информации в инфраструктурах пространственных данных на тестовом полигоне «Мордовия» (расположен между 53°38' и 55°11' с.ш., 42°11' и 46°45' в.д.), раскрывающих особенности взаимодействия лесостепных и лесных геосистем Приволжской возвышенности и Окско-Донской низменности, для высокоточного картографирования на топологическом уровне целесообразно выделить полигоны (рис. 3).

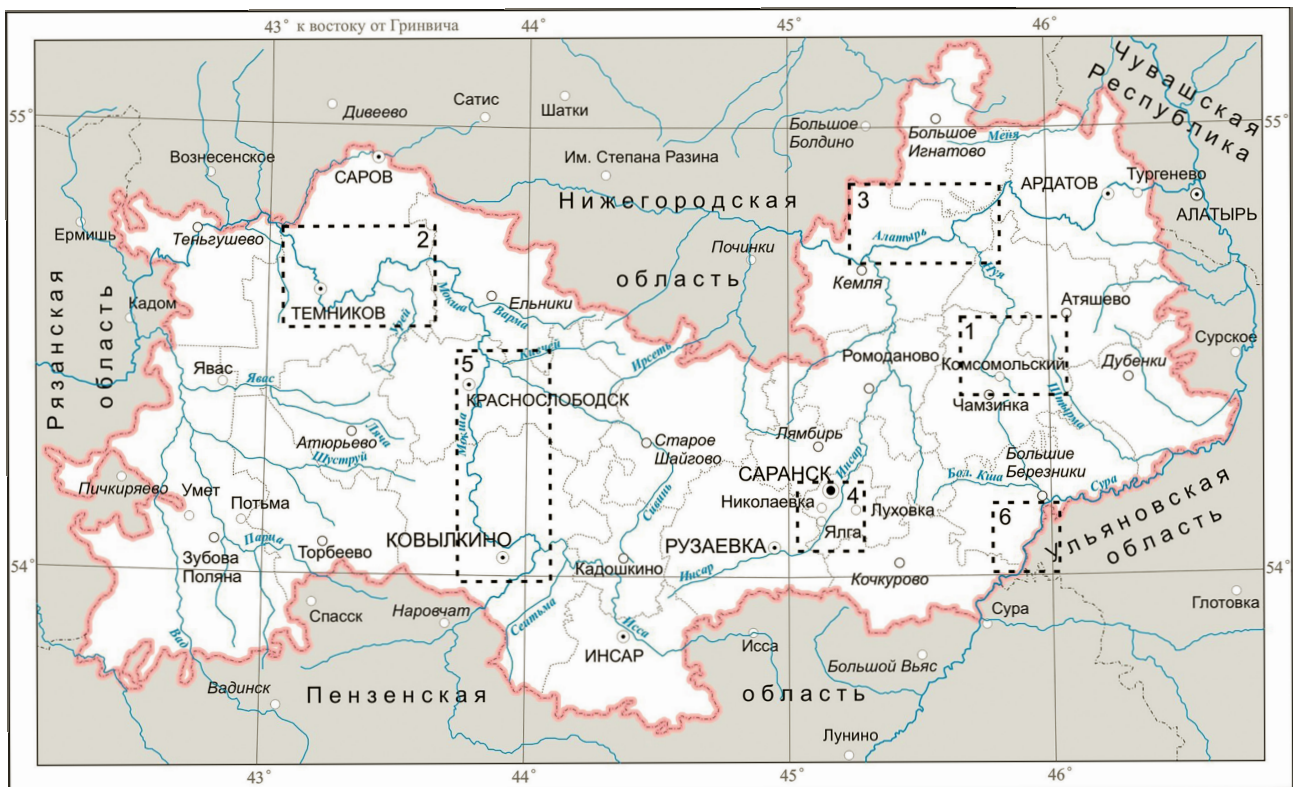


Рис. 3. Тестовые полигоны Республики Мордовия
[Fig. 3. Types of geosystems of the Republic of Mordovia]

1. Полигон «Нуя» (координаты центра: 54°28' с.ш., 45°54' в.д.) отражает взаимодействие геосистем возвышенной эрозионно-денудационной и эрозионно-аккумулятивной равнин плиоцен-плейстоценового возраста. Ключевыми аспектами разработки методов синтетического картографирования являются: диагностика состояния и взаимодействия лесных и лугово-степных геосистем для ландшафтного планирования сельскохозяйственных, лесохозяйственных и горно-технических систем; обоснование зоны экологического равновесия республиканского значения, обеспечивающей стабилизацию водного баланса и минимизацию развития плоскостной и линейной эрозии.

2. Полигон «Мокша–Темников» (координаты центра: 54°42' с.ш., 43°24' в.д.) организован в геосистемах лесных ландшафтов водно-ледниковых и древнеаллювиальных равнин. Основные векторы развития методов диагностики состояния метагеосистем для обоснования комплекса мероприятий по организации паломническо-туристического кластера «Санаксарь–Темников» с сохранением устойчивого функционирования области питания каменноугольно-пермского водоносного комплекса, уникальных экосистем «Заповедной Мордовии».

3. Полигон «Алатырь–Смольный» (координаты центра: 54°48' с.ш., 45°29' в.д.) создан в древней ложбине стока ледниковых вод с хвой-

ными и хвойно-широколиственными лесами; основными объектами исследований являются картографирование экзогеодинамических процессов в зонах линейных тектонических разломов; планирование рекреационных систем.

4. Полигон «*Саранск*» (координаты центра: 54°11' с.ш., 45°11' в.д.) организован для комплексного исследования метагеосистем с целью диагностики развития геоэкологических процессов и ландшафтного планирования градостроительного освоения геосистем.

5. Полигон «*Мокша*» (координаты центра: 54°13' с.ш., 44°02' в.д.) создан для комплексного исследования склоновых и пойменных экзогеодинамических процессов в зонах геотехнических систем.

6. Полигон «*Инерка*» (координаты центра: 54°03' с.ш., 45°53' в.д.) отражает взаимодействие парагенетических систем лесостепных геосистем эрозионно-денудационной равнины и интразональных лесных ландшафтов долины реки Суры. Приоритетной геоэкологической проблемой выступает оптимизация туристско-рекреационного освоения памятника природы республиканского значения «Озеро Инерка».

Анализ пространственных данных с построением синтетической карты геосистем. В последние несколько лет значимое место в области анализа пространственных данных занимает концепция глубокого машинного обучения. Многослойные нейросетевые модели основаны в том числе на сверточных нейросетевых компонентах. Достижение высокой точности интерпретации в данном случае возможно за счет извлечения сложных иерархических признаков и нелинейных зависимостей из пространственно распределенной информации. Однако использование емких глубоких нейронных сетей наталкивается на ряд препятствий, существенно затрудняющих их внедрение. Во-первых, такие модели способны эффективно обучаться лишь на больших наборах размеченных пространственных данных, формирование которых требует серьезных временных и экономических затрат, в том числе значительных объемов полевых исследований и многих часов постобработки. Во-вторых, глубокие нейросетевые модели не являются панацеей с точки зрения выбора инструмента интерпретации пространственных данных: представляя собой черный ящик, они могут быть подвержены проблеме переобучения, плохого обобщения информации и слабой интерпретируемости. Наконец, процесс

обучения глубоких сверточных нейросетевых моделей предъявляет высокие требования к аппаратному обеспечению: проведение экспериментальных исследований по тонкой настройке модели может затянуться без использования дорогостоящих графических процессоров.

Решение обозначенных проблем возможно при одновременной проработке двух направлений: проектировании методов и алгоритмов интеграции и извлечения информативных территориальных признаков пониженной размерности и внедрении облегченных моделей их интерпретации. Важное направление в данной области представляет применение ансамблей классификаторов, выстраиваемых на базе методологии Ensemble Learning, объединяющих в систему различные модели и позволяющих достичь повышения точности и устойчивости моделей машинного обучения [5].

Алгоритм проектирования моноклассификаторов ансамбля. Данные о территориальных системах могут быть многомерными, включая информацию о спектральных свойствах поверхности Земли, особенностях ее пространственной организации, иную информацию атрибутивного, пространственного и временного характера. Набор признаков локального территориального объекта (в том числе пространственных, спектральных и радиометрических характеристик) может быть сформирован на основе материалов космической съемки или иных источников, при этом пространственную область можно классифицировать на основе попиксельного анализа либо путем выделения признаков из фрагментов области разного размера. Совокупность признаков локального объекта, которые сами по себе могут быть представлены в форме тензоров, определяет базовый уровень создаваемой геопрограммной модели местности.

Под геосистемной моделью территории будем понимать данные, характеризующие вмещающие иерархические геосистемы. В соответствии с геосистемным подходом вмещающая геосистема оказывает существенное влияние на свойства иерархически подчиненных образований. Полезным источником информации о ней являются данные дистанционного зондирования Земли. Процесс получения данных различного уровня иерархии потенциально подлежит автоматизации посредством интеграции с сервисами инфраструктуры пространственных данных и сторонними поставщиками пространственной информации.

Оптимизационный алгоритм построения модели классификации метагеосистем включает следующую последовательность этапов: 1) разработка системы требований к модели анализа данных, определения реестра анализируемых данных, качественных и количественных требований к результату функционирования; 2) проектирование базовой архитектуры модели с декомпозицией компонентов верхнего уровня на линейные и ветвящиеся структуры, определение формы входных и выходных сигналов, допустимых показателей качества системы; 3) итерационная оптимизация модели в соответствии с правилом «меньшее лучше большего», для достижения эффекта снижения требований к вычислительным ресурсам и создания предпосылок для решения проблемы переобучения; 4) эвристическое конфигурирование гиперпараметров модели с проектированием метрик точности, алгоритмов оптимизации, функций потерь и количества эпох обучения; 5) анализ вероятностно обусловленного процесса обучения модели посредством мониторинга зависимости математического ожидания и стандартного отклонения меры эффективности классификации от эпохи обучения; 6) оценка качества модели на основе экспериментально выстраиваемых матриц ошибок и расчета метрик точности и ошибочности для вывода о соответствии результата предъявляемым на первом этапе требованиям. Предлагаемая цепочка действий приводит к получению экземпляра модели, а весь процесс поиска наиболее эффективного результата формализуется в виде дерева, терминальные узлы которого соответствуют сформированным моделям анализа пространственных данных.

Ансамбль моделей машинного обучения формируется на основе отдельных обученных моделей (моноклассификаторов) и метаклассификатора, осуществляющего принятие результирующего решения при решении задачи классификации территориального объекта для определения его принадлежности к классу метагеосистем. Отличие моноклассификаторов ансамбля может заключаться как в их архитектурной и топологической организации, так и в данных, использованных для обучения модели. При классификации территориального объекта моноклассификаторы ансамбля формируют вектор гипотез относительно принадлежности данного объекта определенному классу метагеосистем из сформированной номенклатуры классов. При этом каждый обученный моноклассификатор при на-

коплении опыта решения задачи классификации метагеосистем возвращает набор байесовских вероятностей, определяющих степень уверенности моноклассификатора в истинности факта принадлежности территории конкретному классу метагеосистем. Решение о его истинности может быть вынесено путем выбора класса, для которого рассчитанная байесовская вероятность максимальна.

Результирующая гипотеза об отнесении территории к конкретному классу метагеосистем выносится метаклассификатором ансамбля. В этом случае результирующее решение на основе выходных данных моделей глубокого обучения целесообразно принять на основе взвешенного голосования, обобщенное представление которого имеет следующий вид:

$$Y_{Ej} = \operatorname{argmax}_k \left(\sum_{i=1}^N \varphi(M_i, k) \cdot \Psi(p_{jik}) \right) \quad (1)$$

В данной формуле параметр $\varphi(M_i, k)$ – весовой коэффициент, представляющий собой меру эффективности моноклассификатора M_i в детекции метагеосистем класса k . Функция Ψ определяет преобразование вида $\varphi: M_i \rightarrow \Lambda_{ik}$, при котором весовой коэффициент и мера эффективности Λ_{ik} определяют путем математических преобразований экспериментально полученных данных матрицы ошибок M_i моноклассификатора M_i .

Предлагается следующий алгоритм расчета меры эффективности Λ_{ik} .

1. Построение матрицы ошибок M_i для каждого моноклассификатора M_i системы.

2. Вычисление абсолютных метрик точности для классификатора M_i при определении метагеосистем класса k : попаданий (TP_{ik}), верных отклонений (TN_{ik}), ошибок I (FP_{ik}) и II (FN_{ik}) типов.

3. Расчет относительной метрики R_{ik} , определяющей точности классификации метагеосистем класса k , позволяющей осуществить интегральную оценку полученной матрицы ошибок M_i числом в интервале $[0; 1]$. Так, оценка F_β комплексно учитывает показатели точности (*precision*) и полноты (*recall*), а, следовательно, ошибки I и II типов, а также число верных попаданий моноклассификатора. Кроме этого, метрика настраивается посредством конфигурирования параметра β , позволяющего расставить акценты влияния точности и полноты на результат.

$$R_{ik} = F_{i\beta k} = (1 + \beta^2) \cdot \frac{\text{precision}_{ik} \cdot \text{recall}_{ik}}{(\beta^2 \cdot \text{precision}_{ik}) + \text{recall}_{ik}} = \frac{(1 + \beta^2) \cdot TP_{ik}}{(1 + \beta^2) \cdot TP_{ik} + \beta^2 \cdot FN_{ik} + FP_{ik}} \quad (2)$$

Ключ к расчету информативных характеристик территориальных метагеосистем, анализируемых посредством моделей и ансамблей машинного обучения, дает представление о георазнообразии, определяющемся как разнообразие литогенной основы ландшафтов, особенностей почв и растительности и протекающих в них процессов и неявно характеризующем гидрологические и климатические процессы. Ландшафтное разнообразие при этом представляет собой более комплексное понятие и определяет системную организацию пространственно-временных элементов разного уровня: классов, групп, типов, родов и видов геосистем. Ландшафтное разнообразие можно определить как форму абстракции реального мира, в рамках которой свойства территориальных объектов и процессов определены численными переменными или качественными понятиями и могут быть систематизированы по классам.

К первому классу признаков отнесем динамические дескрипторы территории, рассчитываемые на основе данных дистанционного мониторинга (Ддз) и цифровых ландшафтных карт (Длк). Параметры Ддз, которые могут быть рассчитаны на основе изображения территории в определенных спектральных диапазонах (например, в видимом спектре). Предъявим важное ограничение к рассчитываемым параметрам: они должны обладать свойством наглядности – картограммы, построенные на их основе, должны быть информативными для специалистов в области анализа данных наук о Земле.

В ходе эксперимента на основе космической съемки территории были рассчитаны следующие дескрипторы группы Ддз.

1. Ландшафтные метрики неоднородности, основанные на расчете информационной энтропии территории и разброса (характеризующего изменение спектральной яркости относительно среднего значения). Гистограмма оттенков, представляющая собой показатель, выстраиваемый по принципу гистограмм и характеризующий распределение количества пикселей изображения определенных оттенков по группам.

2. Метрики интенсивности, основанные на вычислении инвариантных к шумам и нежелательным искажениям достоверных и стабильных параметров изображения территории. Так,

цветовой момент характеризует распределение спектральной яркости территориального участка и определяется как набор математического ожидания, дисперсии и асимметрии яркости атомарного участка территориальной системы в определенном спектральном диапазоне.

Цифровые ландшафтные карты, генерируемые в региональных и федеральных географических информационных системах, также представляют собой весомый источник информации, позволяющий повысить точность классификации метагеосистем. Несмотря на то, что масштаб таких карт часто меньше масштаба классифицируемых участков, они интегрируют в себе значительное количество информации о вмещающих геосистемах. Получение данных об участке ландшафтной карты может быть осуществлено за счет интеграции модуля подготовки данных с программными API-интерфейсами ГИС-системы посредством использования атрибутивных сведений о координатах анализируемого участка. В ходе эксперимента дескрипторы группы Длк рассчитывались на основе данных региональной ГИС «Мордовия» вычислением метрик, аналогичных тем, что использованы для определения дескрипторов группы Ддз.

Данные метрики представляют собой различные способы численной оценки ландшафтного разнообразия территории и могут быть рассчитаны для участков разного масштаба (геосистем различного иерархического уровня) на основе различных данных (например, изображение территории в различных спектральных полосах).

Важнейший источник информации об инвариантных свойствах территории представляют собой цифровые модели рельефа (ЦМР), которые могут консолидироваться из разных источников. Важным источником информации о рельефе территории часто становятся региональные геоинформационные системы. На основе анализа цифровой модели рельефа территории были рассчитаны следующие дескрипторы группы Дцмр.

1. Метрики крутизны и экспозиции склонов территории, характеризующие поверхностный сток, эрозию, количество получаемой солнечной энергии, рассчитываются на основе определения параметров максимума, минимума, математического ожидания и отклонения перепада высот

между соседними точками ЦМР. Значение крутизны при этом можно вычислить на основе значений карты высот. Экспозиция склонов – морфометрический параметр, отражающий ориентацию исследуемого участка к потоку солнечных лучей и вероятное направление стока воды.

2. Статистические характеристики (минимум, максимум, математическое ожидание, среднеквадратичное отклонение) абсолютных и относительных значений высот ЦМР территории также несут информацию, позволяющую в итоге снизить размерность анализируемых данных.

Формулы расчета обозначенных территориальных дескрипторов представлены в таблице 1. Любое снижение размерности анализируемых данных приводит к потере определенного количества информации, однако в том случае, если результирующий вектор параметров позволяет осуществлять идентификацию территории с допустимой точностью, уменьшение размерности дает возможность подойти к использованию менее глубоких и более устойчивых к проблеме переобучения моделей машинного обучения.

Системный анализ территориальных дескрипторов дает существенный прирост в точности классификации метагеосистем: учет дескрипторов, рассчитанных на основе данных космической съемки территории (группа D_{дзз}), позволил достичь точности 76 %. Вовлечение дескрипторов рельефа (группа D_{цмр}) увеличивает точность на 3 %, а метрик, рассчитанных на основе ландшафтных карт (группа D_{лк}), – на 11 %. Одновременный учет дескрипторов всех категорий приводит к увеличению точности классификации почти на 12 %.

Таким образом, при решении конкретных задач нейросетевые модели малой емкости при условии учета территориальных дескрипторов показывают характеристики точности, сравнимые с показателями сверточных моделей, обученных на многомерных пространственных данных. Важным преимуществом нейросетевых моделей малой емкости являются устойчивая работа в условиях дефицита размеченных данных. Уменьшение доли обучающей выборки до 20 % не привело к весомому снижению точности классификации метагеосистем (в то время как точность сверточной нейросетевой модели упала более чем на 10 %). Это обусловлено высокой информативной нагрузкой выделяемых территориальных дескрипторов, их инвариантностью к изменению исследуемого полигона, а также хорошей способностью модели к обобщению анализируемых данных.

Объединение искусственных нейронных сетей в ансамбль согласно описанной методике с последующим расчетом F_1 -метрики и других параметров эффективности на основе матрицы ошибок показало, что увеличение емкости моделей не приводит к однозначному улучшению результата, так как более мощные модели могут быть более неустойчивы к переобучению, а также требуют большего объема размеченных данных для обучения. При объединении моделей в ансамбль результирующая гипотеза стала применяться на основе взвешенного голосования на основе меры эффективности, что позволило избежать грубых ошибок в классификации, свойственных для каждого классификатора в отдельности. При этом ансамбль лишь незначительно проигрывает в точности отдельным классификаторам системы, сохраняя общую устойчивость к ошибкам I и II рода, свойственных для отдельного классификатора, при определении объектов конкретного класса территории.

ЗАКЛЮЧЕНИЕ

Иерархическое структурирование геосистем оптимизирует диагностику ведущих факторов взаимодействия физико-географических факторов, закономерностей пространственно-временной смены их состояний, направленности развития процессов метаболизма и трансформации вещества и энергии. Общая схема составления цифровой карты метагеосистем региональной ГИС реализуется на основе решения задач интеграции пространственных данных и их систематизации на основе построения иерархии геосистем и проектно-ориентированного применения автоматизированных вычислительных процедур.

Использование ансамблей, выстраиваемых по предложенной методике, позволяет проводить оперативный автоматизированный анализ пространственных данных для решения задачи тематического картографирования метагеосистем и природных процессов. Объединение моделей в ансамбль на основе предложенной архитектуры метаклассификатора позволяет повысить устойчивость анализирующей системы: точность решений, принимаемых ансамблем, имеет тенденцию стремиться к точности наиболее эффективного моноклассификатора системы. Ошибочность системы в большинстве случаев не превышает ошибочность наиболее эффективного классификатора, при этом исключается вероятность грубых систематических ошибок, допускаемых отдельными моноклассификаторами. Формирование ме-

Таблица 1

Формулы расчета дескрипторов территории
 [Table 1. Territory descriptor calculation formulas]

Класс / Class	Тип / Type	Название / Title	Формула расчета / Calculation formula	Параметры / Parameters
Динамические дескрипторы, рассчитываемые на основе данных дистанционного мониторинга и цифровых карт (D _{лк} , D _{дзз}).	Метрики неоднородности	Энтропия территории	$E = \sum_{i=1}^R \frac{n_i}{S} \log \left(\frac{n_i}{S} \right)$	R – радиометрическое разрешение; n _i – количество атомарных единиц спектральной яркости i на территории; S – площадь анализируемой территории.
		Разброс значений	$\Delta = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$	N – количество атомарных участков в анализируемой территории; x _i – среднее значение спектральной яркости атомарного участка; \bar{x} – среднее значение спектральной яркости анализируемой территории.
		Гистограмма яркости	$P = \begin{cases} h_a = \sum_{i=1}^N 1, \text{ если} \\ x_i \geq \min(x_i) + (a-1)\sigma \\ x_i < \min(x_i) + a\sigma \\ \sigma = \frac{(\max(X) - \min(X))}{A} \end{cases}$	X – множество значений спектральной яркости для атомарных участков территории; A – размер шага гистограммы.
Статические дескрипторы, рассчитываемые на основе цифровой модели рельефа (D _{цмр}).	Метрики интенсивности	Математическое ожидание	$\bar{I} = \frac{1}{N} \sum_{j=1}^N c_j$	c _j – яркость -го участка в определенном спектре
		Дисперсия	$D = \sqrt{\frac{1}{N} \sum_{j=1}^N (c_j - \bar{I})^2}$	
		Асимметрия	$A = \sqrt[3]{\frac{1}{N} \sum_{j=1}^N (c_j - \bar{I})^3}$	
		Крутизна	$\alpha = \arctg \left(\frac{1}{2\sigma} \sqrt{(h_E - h_W)^2 + (h_N - h_S)^2} \right)$	σ – масштабирующий коэффициент, h _{E,W,N,S} – высота восточной, западной, северной, южной точек
		Экспозиция	$\beta = -\arctg \left(\frac{h_E - h_W}{h_N - h_S} \right)$	
		Статистические параметры	$\gamma = \text{stat}(H)$	H – матрица высот, stat – статистическая функция

так классификатора по предложенному алгоритму представляет собой возможность добавить элемент предсказуемости и контроля в использование нейронных сетей, традиционно представляющих собой «черный ящик».

Совокупный анализ дескрипторов территории, интегрируемых на основе данных из разных источников, дает существенный прирост точности классификации метагеосистем. В рамках эксперимента, представленного в статье, учет предложенной системы дескрипторов, рассчитанных на основе данных космической съемки территории, цифровой модели рельефа и электронной ландшафтной карты, позволил достичь точности 89%, что значительно больше этого параметра для сверточной нейросетевой модели. При этом анализ дескрипторов рельефа увеличивает точность на 3%, а метрику, рассчитанных на основе ландшафтных карт, – на 11%. Картограммы представленных дескрипторов хорошо интерпретируются специалистами в области анализа данных в науках о Земле.

Расчет и консолидация территориальных дескрипторов, предложенных нами, приводят одновременно к двум последствиям: снижению размерности анализируемых данных (положительный эффект) и неизбежной утрате некоторого количества информации об анализируемой территории (отрицательное влияние). Важно найти баланс между

двумя обозначенными положениями, с тем чтобы максимально облегчить емкость модели машинного обучения, повысить ее устойчивость к переобучению, не допустить значительного снижения точности классификации в рамках конкретной решаемой задачи.

СПИСОК ЛИТЕРАТУРЫ

1. Николаев В. А. *Ландшафтоведение*. Москва: Издательство Московского университета, 2000. 94 с.
2. Сочава Б. В. *Введение в учение о геосистемах*. Новосибирск: Наука. Сибирское отделение, 1978. 319 с.
3. Черкашин А. К. Предмет географических исследований: метагеографический подход // *Известия Русского географического общества*, 2022, т. 154, № 2, с. 3-21.
4. Improving the Efficiency of Deep Learning Methods in Remote Sensing Data Analysis: Geosystem Approach / S.A. Yamashkin, A.A. Yamashkin, V.V. Zanozin et. al. // *IEEE Access*, 2020, vol. 8, pp. 179516-179529.
5. Polikar R. Ensemble based systems in decision making // *IEEE Circuits and systems magazine*, 2006, vol. 6, no. 3, pp. 21-45.
6. Sagi O., Rokach L. Ensemble learning: A survey // *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2018, vol. 8, no. 4, pp. 1249.

Конфликт интересов: Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Поступила в редакцию 08.06.2022

Принята к публикации 30.11.2022

Test Polygons for Geosystems State Diagnostics and Development of Earth Remote Sensing Data Interpreting Techniques

A. A. Yamashkin , S. A. Yamashkin,
M. R. Baichurin, A. A. Tokarev, I. S. Lyamzina

National Research Mordovian State University, Russian Federation
(68, Bolshevistskaya str., Saransk, 430005)

Abstract: The study presented in the article is aimed at developing methods and algorithms for system analysis of test site data in order to diagnose the state of geosystems and develop methods for interpreting Earth remote sensing data.

Materials and methods. The development of methods and algorithms for the analysis and integration of spatial information was based on the analysis of a system of test sites that reveal the features of the interaction between the forest-steppe and forest geosystems of the Volga Upland and the Oka-Don Lowland. The general scheme of the process of compiling a digital map of metageosystems is implemented by solving the following tasks: collecting and preparing a system of thematic maps and databases; systematization of information with the construction of a hierarchy of geosystems; ensemble analysis of multi-zone satellite images with the construction of a synthetic map of geosystems; evaluation of simulation results; obtaining and practical use of spatial information.

Results and discussion. Combining models into an ensemble based on the proposed architecture of the metaclassifier makes it possible to increase the stability of the analyzing system: the accuracy of decisions made by the ensemble tends to tend to the accuracy of the most efficient monoclclassifier of the system. Systematic analysis of territory descriptors integrated on the basis of data from different sources provides a significant increase in the accuracy of metageosystem classification. It is important that the cartograms of the presented descriptors are well interpreted by specialists in the field of data analysis in the geosciences.

Conclusion. The use of ensembles built according to the methodology proposed in the article makes it possible to carry out an operational automated analysis of spatial data to solve the problem of thematic mapping of metageosystems and natural processes. The calculation and consolidation of territorial descriptors makes it possible to reduce the dimensionality of the analyzed data, to facilitate the allowable capacity of the machine learning model, to increase its resistance to overfitting, and to prevent a significant decrease in the classification accuracy within the framework of a specific problem being solved.

Key words: spring water samples, inventory, priority pollutants, non-centralized water sources, drinking water quality.


Funding: The study was supported by the Russian Science Foundation grant No. 22-27-00651, <https://rscf.ru/en/project/22-27-00651/>

For citation: Yamashkin A. A., Yamashkin S. A., Baichurin M. R., Tokarev A. A., Lyamzina I. S. Test Polygons for Geosystems State Diagnostics and Development of Earth Remote Sensing Data Interpreting Techniques. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Seria: Geografia. Geoekologia*, 2022, no. 4, pp. 4-18 (In Russ.) DOI: <https://doi.org/10.17308/geo/1609-0683/2022/4/4-18>

REFERENCES

1. Nikolaev V. A. *Landschaftovedenie* [Landscape Science]. Moscow: Izdatel'stvo Moskovskogo universiteta, 2000. 94 p. (In Russ.)
2. Sochava B. V. *Vvedenie v uchenie o geosistemah* [Introduction to the study of geosystems]. Novosibirsk: Nauka. Sibirskoe otdelenie, 1978. 319 p. (In Russ.)
3. Cherkashin A. K. *Predmet geograficheskikh issledovaniy: metateoreticheskij podhod* [Subject of geographical research: metatheoretical approach]. *Izvestiya Russkogo geograficheskogo obshchestva*, 2022, vol. 154(2), pp. 3-21. (In Russ.)
4. Improving the Efficiency of Deep Learning Methods in Remote Sensing Data Analysis: Geosystem Approach

© Yamashkin A. A., Yamashkin S. A., Baichurin M. R., Tokarev A. A., Lyamzina I. S., 2022

 Anatoliy A. Yamashkin, e-mail: yamashkin56@mail.ru



The content is available under Creative Commons Attribution 4.0 License.

/ S.A. Yamashkin, A.A. Yamashkin, V.V. Zanozin et. al. *IEEE Access*, 2020, vol. 8, pp. 179516-179529.

5. Polikar R. Ensemble based systems in decision making. *IEEE Circuits and systems magazine*, 2006, vol. 6, no. 3, pp. 21-45.

6. Sagi O., Rokach L. Ensemble learning: A survey. *Wiley Interdisciplinary Re-views: Data Mining and Knowledge Discovery*, 2018, vol. 8, no. 4, pp. 1249.

Ямашкин Анатолий Александрович

доктор географических наук, профессор, декан географического факультета национального исследовательского Мордовского государственного университета, г. Саранск, Российская Федерация, ORCID: 0000-0001-9995-8371, e-mail: yamashkin56@mail.ru

Ямашкин Станислав Анатольевич

доцент кафедры автоматизированных систем обработки информации и управления национального исследовательского Мордовского государственного университета, г. Саранск, Российская Федерация, ORCID: 0000-0002-7574-0981, e-mail: yamashkin56@mail.ru

Байчурин Марат Радикович

аспирант географического факультета национального исследовательского Мордовского государственного университета, г. Саранск, Российская Федерация, e-mail: baichurin.marat@yandex.ru

Токарев Александр Александрович

аспирант географического факультета национального исследовательского Мордовского государственного университета, г. Саранск, Российская Федерация, e-mail: killen00@yandex.ru

Лямзина Инна Сергеевна

аспирант географического факультета национального исследовательского Мордовского государственного университета, г. Саранск, Российская Федерация, e-mail: innapooh@rambler.ru

Conflict of interests: The authors declare no information of obvious and potential conflicts of interest related to the publication of this article.

Received: 08.06.2022

Accepted: 30.11.2022

Anatoliy A. Yamashkin

Dr. (Geogr.) Sci., Professor, Dean of the Faculty of Geography, National Research Mordovian State University, Saransk, Russian Federation, ORCID: 0000-0001-9995-8371, e-mail: yamashkin56@mail.ru

Stanislav A. Yamashkin

Associate Professor of the Department of Automated Information Processing and Control Systems, National Research Mordovian State University, Saransk, Russian Federation, ORCID: 0000-0002-7574-0981, e-mail: yamashkinsa@mail.ru

Marat R. Baichurin

Postgraduate student of the Faculty of Geography, National Research Mordovian State University, Saransk, Russian Federation, e-mail: baichurin.marat@yandex.ru

Aleksandr A. Tokarev

Postgraduate student of the Faculty of Geography, National Research Mordovian State University, Saransk, Russian Federation, e-mail: killen00@yandex.ru

Inna S. Lyamzina

Postgraduate student of the Faculty of Geography, National Research Mordovian State University, Saransk, Russian Federation, e-mail: innapooh@rambler.ru