

**Справочник по вопросам
использования данных наблюдения
Земли для целей официальной
статистики**

**Отчет Целевой группы по
спутниковым изображениям и
геопространственным данным**

5-е декабря 2017 г.

ООН

Австралийское бюро статистики

Технологический университет Квинсленда, Австралия

Правительство Квинсленда, Австралия

Государственное объединение научных и прикладных исследований, Австралия

Европейская комиссия – Генеральный Директорат Евростата

Национальный институт статистики и географии, Мексика

Статистическое управление Канады

Основные положения

Данная публикация, подготовленная членами Целевой группы Организации Объединенных Наций по спутниковым изображениям и геопространственным данным, представляет собой справочник для национальных статистических управлений, рассматривающих возможность использования данных спутниковых снимков (известных как данные наблюдения Земли) для целей официальной статистики. Справочник был подготовлен в качестве вклада в доклад Глобальной рабочей группы по вопросам использования больших данных для подготовки официальной статистики Статистической комиссии Организации Объединенных Наций за 2017 год.

Справочник содержит краткое введение в использование данных наблюдений Земли (НЗ) для официальной статистики, типы доступных источников и методологии получения статистики на основе данных этого типа. В нем также обобщаются результаты четырех пилотных проектов, подготовленных членами Целевой группы; Австралийским бюро статистики (ABS), Национальным институтом статистики и географии Мексики (INEGI), Национальным административным департаментом статистики Колумбии (DANE) и компанией Google. Дополнительный материал включает обзор литературы по методологии и полные отчеты о пилотных проектах.

В заключительной главе представлены руководящие принципы для национальных статистических управлений, которые следует учитывать и на которые рекомендуют ссылаться при рассмотрении вопроса о том, следует ли внедрять источники данных наблюдений Земли в процесс статистического производства, а также вопросы, которые желательно учитывать в процессе внедрения и получения результатов. Эти руководящие принципы не носят директивного характера, а скорее призваны побудить Национальные статистические управлении критически рассмотреть свои собственные деловые потребности и применять рекомендации в соответствии с обстоятельствами. По мере роста международного опыта использования этих данных в статистических целях национальные статистические управлении могут опираться на руководящие принципы, представленные в этой публикации, с учетом собственного опыта и знаний.

Предисловие

Официальные статистики десятилетиями использовали различные источники данных при производстве официальной статистики, включая «созданные» источники данных, такие как переписи и обследования, и «найденные» источники данных, такие как административные и транзакционные данные.

В результате все более широкого взаимодействия граждан с цифровыми технологиями и расширения возможностей этих технологий по предоставлению цифровых следов появились новые источники данных, которые становятся все более доступными для официальных статистиков. К таким источникам относятся данные сенсорных сетей и устройств слежения, например спутников и мобильных телефонов.

Статистические агентства во всем мире очень заинтересованы в изучении жизнеспособности использования данных спутниковых изображений или, в более широком смысле, данных наблюдения Земли (НЗ), для улучшения официальной статистики по широкому кругу вопросов, охватывающих сельское хозяйство, окружающую среду, деловую активность и транспорт.

Данные НЗ обладают значительным потенциалом для предоставления более своевременной статистической информации, сокращения частоты обследований, снижения нагрузки на респондентов и сопутствующих расходов, предоставления данных на более дезагрегированном уровне для принятия обоснованных решений и предоставления новых статистических данных и статистических знаний.

Данные НЗ также могут способствовать мониторингу Целей в области устойчивого развития (ЦУР) за счет повышения своевременности и актуальности показателей без ущерба для их беспристрастности и методологической обоснованности.

В этом справочнике представлено краткое введение в использование данных НЗ для официальной статистики, типы доступных источников, примеры использования данных НЗ для составления показателей ЦУР и методологии получения статистики на основе данных этого типа. В нем также обобщаются результаты четырех экспериментальных проектов, подготовленных членами Целевой группы ООН по спутниковым снимкам и геопространственным данным для изучения возможности использования данных о наблюдениях для официальной статистики, а также другие соответствующие и полезные тематические исследования. Дополнительный материал включает обзор литературы по методологии.

В заключительной главе представлены руководящие принципы для национальных статистических управлений, которые следует учитывать и на которые следует ссылаться при рассмотрении вопроса об использовании данных НЗ при подготовке официальной статистики, а также вопросы, которые следует учитывать в процессе реализации и получения результатов.

Создание этого справочника было бы невозможно без напряженной работы и значительного вклада авторов каждой из глав. Хотелось бы выразить признательность д-ру Ханнесу И. Рейтеру (Евростат), д-ру Арнольду Деккеру, Флоре Керблат, д-ру Алексу Хелду и д-ру Роберту Вудкуку (Государственное объединение научных и прикладных исследований), Алексис Макинтайр (Геонаука Австралии), профессору Керри Менгерсен и Джеймсу Макбруму (Технологический университет Квинсленда), Сандре Янет Родрикес (Национальный административный департамент статистики Колумбии) и Хасинте Холлоуэй (Австралийское бюро статистики). Также хотелось бы поблагодарить авторов пилотных отчетов Целевой группы: Патрика Дунагана (Google), Роналда Окампо и Хосе Де Ла Торре (Национальный институт статистики и географии Мексики), Дженнифер Марли, Райана Дефина, Кейт Трегер, Даниэля Элазара, Анура Амарасингхе и Гарет Биггс (Статистическое бюро Австралии). Мы также благодарим Бена Фитцпатрика и Бриджит Колин (Технологический университет Квинсленда) за их вклад в тематические исследования методов в главе 4. Отдельно благодарим Статистическое управление Канады за предоставленную

нам возможность включить их работу в качестве примера использования на практике данных НЗ для официальной статистики.

Получение комментариев и отзывов от экспертов в области науки о Земле, статистической методологии, сельского хозяйства и геопространственных областей было важным для создания этого руководства. В этой связи хотелось бы поблагодарить Сибилл МакКаун, Даниэля Элазара, Мартина Брэди и Томаса Уолтера (Статистическое бюро Австралии), Алексис Макинтайр (Геонаука Австралии), д-ра Андриса Потгитера (Университет Квинсленда), д-ра Мэттью Прингла и д-ра Майкла Шмидта (Департамент Квинсленда по наукам, информационным технологиям и инновациям), доктора Джорджа Кончедда и Жака Делинса (ФАО ООН) и Хавьера Гальяго (Объединенный исследовательский центр (JRC) Европейской комиссии) за их комментарии и предложения по улучшению содержания справочника.

Д-р Сю-Мин Там – Председатель Целевой группы ООН по спутниковым изображениям и геопространственным данным

1. Введение

Дистанционное зондирование — это технология, направленная на наблюдение и изучение систем Земли и их динамики. Данные, относящиеся к дистанционному зондированию, в частности спутниковые снимки, имеют цель наблюдать и изучать Землю из космоса; поверхности суши, океанов и атмосферы (ESA, 2016 г.)¹. Цивилизация взаимодействует с природными компонентами земных систем, адаптируя их разнообразными и сложными способами, создавая модифицированные земные системы, пригодные для измерения с помощью наблюдения Земли (НЗ).

В этом смысле то, что делает общество, и его отношения с Землей имеют разные реакции, которые изменяют геопространственные переменные. Они могут быть прямо или косвенно обнаружены в ходе НЗ; поскольку в действительности, социально-экономическая деятельность распределена по большей части нашей планеты. Затем, если национальные статистические управления (НСУ) являются одними из наиболее важных источников базовой социально-экономической информации для своих правительств, то данные и информация НЗ являются ключевым геопространственным источником, который дополняет, улучшает и обновляет официальную статистику. Это означает смену парадигмы с прошлых и нынешних статистических подходов к измерению эффектов социальной жизни на Земле (антропогенных эффектов), ее эволюции, распространения, воздействия и устойчивости. В основном главные проблемы, с которыми сталкивается человечество в том, что сейчас называется Антропоценом.

Данные НЗ и информация, полученная из этих данных, представляют собой частный случай больших геопространственных данных. Большие данные определяются как «любой большой и сложный набор данных, который становится трудно обрабатывать с помощью традиционных приложений для обработки данных», и, следовательно, включают данные НЗ со спутников. Во многих случаях большие данные считаются неструктуризованными данными или данными, которые используются для другой цели, чем предполагал первоначальный создатель данных. В случае данных НЗ, они фактически специально созданы, чтобы быть высокоструктуризованными и измерять определенные аспекты Земли. Сложность данных НЗ заключается в извлечении значимой информации о реальном объекте, значении, состоянии или состоянии из его отраженного или излучаемого электромагнитного сигнала, исходящего от Земли, атмосферы или океана, проходящего через атмосферу и измеренного в космосе, океане или почве.

Спутниковая геологоразведка как источник информации в поддержку многих секторов правительства и промышленности была подтверждена мировыми лидерами в сентябре 2015 года при принятии Повестки дня в области устойчивого развития на период до 2030 года. Они признали исключительную важность «прозрачного и подотчетного расширения соответствующего государственно-частного партнерства для использования вклада, который должен внести широкий спектр данных, включая информацию о Земле и геопространственных данных, при обеспечении национальной ответственности за поддержку и отслеживание прогресса». Принимая Повестку дня в области устойчивого развития на период до 2030 года, мировые лидеры согласились с тем, что система глобальных показателей станет важным инструментом для измерения, мониторинга и отчетности о прогрессе в достижении Целей в области устойчивого развития (ЦУР). На 47-й сессии Статистической комиссии ООН 8 – 11 марта 2016 г. были представлены и обсуждены глобальные показатели, которые помогут НСУ осуществлять мониторинг этих 17 ЦУР. Эффективное использование информации со спутниковых изображений может оказать трансформирующее воздействие на многие из наиболее серьезных проблем человечества, таких как помочь мировым ученым, менеджерам по ресурсам и планированию, а также поможет политикам лучше контролировать и защищать хрупкие экосистемы, обеспечивать устойчивую инфраструктуру, управлять климатическими рисками, повышать качество продовольствия, строить

¹ http://www.esa.int/SPECIALS/Eduspace_EN/SEMF9R3Z2OF_0.html

более устойчивые города, сокращать бедность и повышать эффективность управления, среди прочего.

Глобальная рабочая группа (ГРГ) по большим данным для официальной статистики была создана в 2014 году по итогам 45-го заседания Статистической комиссии ООН. В соответствии со своим кругом ведения ГРГ ООН (UN GWG) обеспечивает стратегическое видение, направление и координацию глобальной программы по большим данным для официальной статистики, в том числе для показателей повестки дня в области устойчивого развития на период до 2030 года. Она также способствует практическому использованию источников больших данных, способствуя наращиванию потенциала, обучению и обмену опытом.

Наконец, ГРГ ООН (UN GWG) содействует коммуникации и пропаганде использования больших данных для целей политики и предлагает советы по укреплению общественного доверия к использованию больших данных в частном секторе.

В этом контексте в данной публикации основное внимание уделяется данным и методам исследования НЗ Целевой группой ГРГ ООН (UN GWG) по спутниковым изображениям и геопространственным данным.

В прошлом освоение НСУ информации о НЗ было медленным. Тем не менее, многие НСУ в настоящее время открывают для себя потенциал использования новых и согласованных источников данных и методологий для интеграции различных «географических» переменных с целью поддержки и информирования официальной статистики. Они могут быть созданы с помощью комбинации геопространственной информации, геологоразведки и других источников больших данных и способны заполнить пробелы в данных, предоставить информацию там, где никогда не проводились измерения, а также улучшить временное и пространственное разрешение данных (например, ежедневные обновления посевных площадей и статистики урожайности).

Такой сдвиг в парадигме от традиционных статистических методов (например, подсчета, измерения людьми) к оценке с помощью датчиков (с соответствующими аспектами достоверности, аппроксимации, неопределенности), симуляции и моделирования не будет простым, поскольку в настоящее время существует устоявшаяся статистическая система со своими преимуществами и недостатками. Для этого потребуются убедительные, статистически обоснованные результаты, полученные в производственном процессе в течение ряда лет, хорошо продуманные аргументы и превосходные научные методы, готовность перераспределять ресурсы в учреждениях, положительный анализ затрат и выгод, кроме того мотивация (например, сокращение затрат/штатных сотрудников) двигаться вперед в быстро меняющемся мире, чтобы адаптироваться к более высоким пространственным и временным разрешениям для решения вопросов политики.

Все это идет рука об руку с включением этих новых методов в Типовую модель производства статистической информации (ТМПСИ). Официальная документация ТМПСИ помогает описать и определить набор бизнес-процессов, необходимых для производства официальной статистики. НСУ предоставляется стандартная структура и согласованная терминология, которые помогают модернизировать процессы статистического производства, а также обмениваться методами и компонентами. Кроме того, ТМПСИ также можно использовать для интеграции стандартов данных и метаданных, в качестве шаблона для документации процессов, с целью гармонизации инфраструктур статистических вычислений и для обеспечения основы для оценки и улучшения качества процессов (ЕЭК ООН, 2013 г.). В настоящее время, согласно опросу ЕЭК ООН (2015 г.), большинству стран все еще необходимо разработать стратегию для больших данных или ТМПСИ с целью соответствующего обращения с этими видами данных.

В противовес этому, производственные процессы, так или иначе похожие на бизнес-процессы или ТМПСИ, действуют за пределами НСУ. Примеры можно найти в различных статистических областях — например, в отношении лесов — Глобальной инициативы по наблюдению за лесами (GFOI) — и более подробно они обсуждаются в главах 2 и 4. Мы выбрали показательную область

сельскохозяйственного производства, в которой две упомянутые ниже организации, например, вносят значительный вклад в Глобальный сельскохозяйственный мониторинг (GEOGLAM).

Министерство сельского хозяйства США ежемесячно выпускает отчет «Оценки мирового предложения и спроса на зерновые» (WASDE) (USDA, 2016 г.) с самыми ранними доступными отчетами, начиная с 1973 года. Продовольственная и сельскохозяйственная организация (ФАО) ООН представляет ежемесячный бюллетень о спросе и предложении зерновых (ФАО, 2016 г.) с дополнительной подробной информацией в ежеквартальной публикации «Перспективы урожая и положение с продовольствием», подготовленном экономистами. Эти отчеты содержат оценки и прогнозы сельскохозяйственного производства, посевных площадей и цен на национальном или даже субнациональном уровне на основе набора данных. Отчеты Министерства сельского хозяйства США четко увязывают свои результаты с аналитическими данными, основанными на информации о погоде и спутниковых данных (например, с пространственным разрешением до 250 м на изображениях MODIS).

Немного другим примером является Европейская комиссия, где бюллетени Мониторинга сельскохозяйственных ресурсов (MARS) (ЕК, 2016 г.) сообщают последние прогнозы роста урожая сельскохозяйственных культур (зерновые, масличные культуры, белковые культуры, сахарная свекла, картофель, рис, подножный корм), включая краткосрочные последствия метеорологических явлений. Эти отчеты поддерживаются с 1992 года мониторингом роста сельскохозяйственных культур и информацией прогнозирования урожайности в режиме, близком к реальному времени, который называется системой прогнозирования урожайности сельскохозяйственных культур MARS (MCYFS). Еще несколько таких систем существуют по всему миру. Последним продуктом, созданным в сотрудничестве со Статистическим управлением Филиппин, является Филиппинская информационная система по рису (PRISM, 2016 г.), целью которой является объединение различных источников данных для определения гораздо большего, чем может охватить традиционный статистический продукт ТМПСИ. PRISM стремится получить такие параметры, как сезонность урожая риса; область; урожайность; ущерб (например, наводнение, ветер, засуха); и факторы, снижающие урожайность (например, болезни, животные-вредители и сорняки), о которых, безусловно, не будет сообщаться со статистической точки зрения, но которые необходимы для поддержки политики, основанной на фактических данных. Статистическому сообществу необходимо определить, с какой целью и с помощью каких инструментов предполагается предоставлять информацию по актуальным и ближайшим вопросам политики, основанным на фактических данных.

Одной из основных возможностей, которую можно рассмотреть в этом контексте, является использование данных НЗ. В настоящее время мы наблюдаем быстрые изменения в сфере НЗ с точки зрения возможностей и доступности. В 1997-1998 гг. из 23 доступных/запрограммированных изображений Landsat-TM для конкретного места за данный год только 2 из соответствующих 4 изображений можно было использовать из-за облачного покрова на северо-востоке Германии (Grenzdörffer, G., 2001 г.). Это полностью контрастирует с наличием засушливых или полузасушливых районов, где, возможно, не было одного или двух изображений. За последние десятилетия потребители данных стали свидетелями расширения глобального охвата изображениями и датчиками (например, MODIS, Sentinels, WorldView, Bella Vista — см. главу 2) вплоть до ежедневных глобальных мозаик (planet labs, 2016 г.). Тем не менее, погода (например, облачность – Wilson & Jetz, 2016 г.) и местоположение (например, Норвегия и время ее солнечного освещения) по-прежнему будут играть роль в уменьшении количества подходящих данных. Этот огромный объем данных вносит изменения в традиционный способ обработки данных НЗ. Спутник EU Copernicus Sentinels будет создавать 3 ПБ данных в год, по всему миру оценки достигнут 1 ЭБ. Сообщество перейдет от локальной обработки одного изображения за определенный день к рабочему процессу, который объединяет а) несколько датчиков (например, SAR, тепловизионный, оптический — см. главу 2) с б) различными другими входными потоками данных в) в разные даты и д) разные разрешения. Таким образом, обработка будет перенесена в высокопроизводительные вычислительные центры обработки данных (например, Google Earth Engine, Planet-API, Национальная вычислительная инфраструктура в Австралии, платформа Digital Globe DGBX) с использованием инструментов обработки больших данных (например, MapReduce, Hipi) по пути перемещения алгоритмов к данным, а не данных к алгоритмам. Следует ожидать большого прироста

информации в сочетании с более сложными неопределенностями, особенно в сочетании с различными источниками данных. Риски и проблемы необходимо снижать (например, Ma и др., 2015 г.; Li и др., 2016 г.); однако выигрыш во временном и пространственном разрешении для решения важных вопросов полиции оправдывает эти риски.

С возможностями и доступностью связана недавняя разработка концепции предоставления Готовых к анализу данных (ARD). Здесь спутниковые данные были обработаны в соответствии с минимальным набором требований и организованы в форме, которая позволяет немедленно добавлять ценность и анализировать их без дополнительных усилий пользователя. Преимуществами являются воспроизводимость, последовательность и более быстрое методологическое развитие. Примером может служить использование набора данных ARD для обнаружения водной поверхности за период в 27 лет, который можно использовать для расчетов площади поверхности суши для целей статистики (Lewis и др., 2016 г.).

И последнее, но не менее важное напутствие руководства по данным и методам НЗ для непосвященных. НЗ могут повторно охватывать большие площади (даже недоступные) с различными спектральными каналами и пространственным разрешением, которые могут быть проанализированы на компьютере для многих приложений (например, базовая карта, растительный покров, оценка урожая, эрозия, качество воды, коралловые рифы, наводнения и прорывы дамб). Некоторые недостатки заключаются в том, что люди – операторы выбирают датчики, калибруют данные и выполняют анализ, но для успеха необходимы специальные знания. НСУ также должны знать, что данные НЗ имеют различные ограничения. НЗ по умолчанию не предоставляют никаких статистических показателей; они предоставляют некоторую пространственную, спектральную и временную информацию, которая может быть связана с индикаторами. Поэтому рекомендуется начинать с малого, чтобы набраться опыта — даже с большими наборами данных, готовых к анализу, — а затем увеличивать масштаб.

Наконец, в этом документе предпринята попытка представить соответствующий обзор данных НЗ (глава 2) и методов (глава 3), продемонстрировать несколько пилотных проектов, собранных по всему миру (глава 4), и дать рекомендации статистикам системы ООН и НСУ о том, как использовать данные НЗ в своих сферах деятельности (Глава 5).

2. Источники данных

2.1 Большие данные и наблюдения Земли

Эта глава направлена на предоставление НСУ руководящей информации для рассмотрения вопроса о том, могут ли они и должны ли они использовать информацию, полученную с помощью спутниковых данных НЗ, в качестве формы больших данных, чтобы иметь отношение, например, к показателям ЦУР ООН или другой соответствующей информации НСУ, и если да, то какие источники будут наиболее актуальными. Эта глава также направлена на то, чтобы пролить свет на некоторые сильные и слабые стороны использования данных НЗ для официальной статистики, что важно, когда от специалистов требуется быстро и надежно предоставлять информацию лицам, принимающим решения.

Существует множество систем для наблюдения Земли (архивные, текущие и запланированные — всего более 100 датчиков), поэтому полезно классифицировать эти датчики по характеристикам их чувствительности. Все это необходимо, чтобы иметь возможность интерпретировать общедоступные списки датчиков НЗ, доступные, например, на веб-сайте CEOS (Комитета по спутникам наблюдения Земли) (www.ceos.org) и, более подробно, на странице базы данных: <http://database.eohandbook.com/>, а также в базе данных OSCAR (Инструмент анализа и обзора возможностей систем наблюдений) ВМО (Всемирной метеорологической организации) <https://www.wmo-sat.info/oscar/satellites>.

Каждая из этих баз данных имеет значение для выбора наиболее подходящих датчиков (см. п.2.1.2 Источники информации о данных наблюдения Земли). На рисунке 1, например, показан выбор для измерения растительного покрова. Это указывает на гарантированную непрерывность данных как минимум до 2030 года в сочетании с проблемой выбора датчика (датчиков) НЗ, наиболее подходящего(их) для этой цели.

VEGETATION COVER

Current and Future Missions



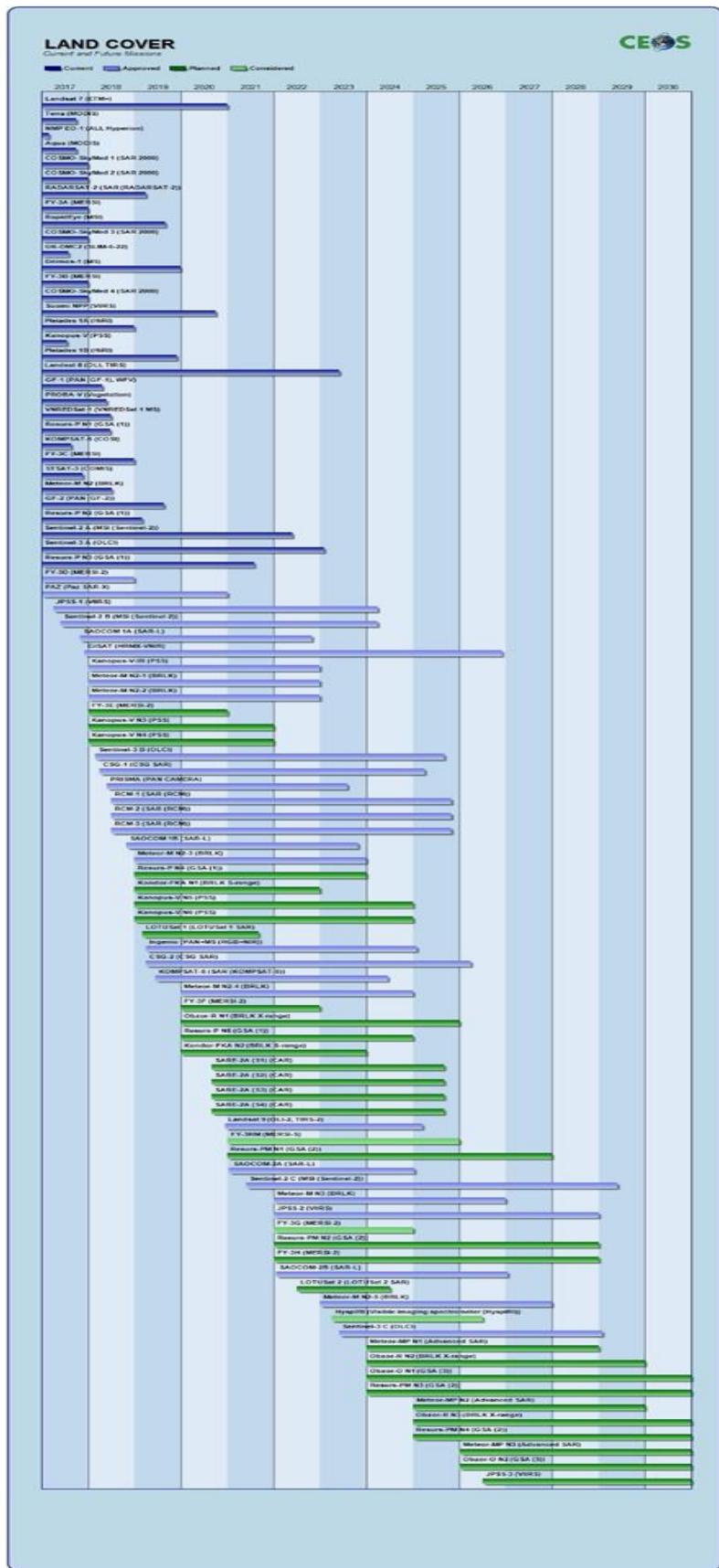


Рисунок 1: Выдержка из базы данных CEOS об измерениях растительного и земного покрова

Источник: <http://database.eohandbook.com/timeline/>

2.2 Две основные системы НЗ: LEO, GEO

Существуют три основных типа систем НЗ в зависимости от их земной орбиты, что влияет на пространственное и временное разрешение; спутники LEO (низкая полярная орбита Земли), МEO (средняя околоземная орбита) и GEO (геостационарные).

Спутники LEO обычно летают на высоте от 400 до 800 км по орбите, которая проходит через оба полюса (с небольшим отклонением, называемым наклоном). Фактически они сканируют Землю на фиксированной орбите со скоростью примерно 28 000 км/ч, в то время как Земля вращается под ними, таким образом создавая глобальную картину в течение некоторого периода времени (от ежедневного до ежемесячного в зависимости от широты полосы захвата, которую они регистрируют, или количества спутников, одновременно находящихся на орбите). Каждый оборот Земли занимает около 90 минут. Эти спутники имеют увеличенный охват в направлении полюсов и уменьшенный охват вокруг экватора (см. рис. 3). Поскольку они находятся относительно близко к Земле, их пространственное разрешение может достигать 30 см в черно-белом (панхроматическом) режиме и около 1 м в цветном или многоспектральных каналах для коммерчески доступных изображений.

Наиболее известными из общедоступных датчиков спутников LEO является Сканирующий спектрорадиометр среднего разрешения MODIS (MODerate-resolution Imaging Spectroradiometer с 1999 г.) и датчики Landsat (с 1984 г.) с пространственным разрешением 250/500/1000 м для MODIS и 30 м для Landsat. 80-метровые данные Landsat доступны с 1973 г. (Примечание: пространственное разрешение, например, 30 м указывает на размер пикселя 30 на 30 м = 900 м²). Недавно запущенный спутник Sentinel-2A (за которым последовали многие другие спутники Sentinel-2B, 2C, 2D и т. д.) имеет пространственное разрешение 10 м (а также некоторые каналы с разрешением 20 м и 60 м).

Примерами общественных благ являются данные Landsat с разрешением 30 м и Sentinel-2 с разрешением 10 м; - предоставление свободно доступных данных со значительным объемом документации и рецензируемой литературы, охватывающей все аспекты датчика для приложений, имеющих отношение к обществу.

Изображения Landsat 4, 5 и 7 состоят из семи спектральных каналов с пространственным разрешением 30 на 30 м для каналов с 1 по 5 и 7. Пространственное разрешение для канала 6 (тепловой инфракрасный канал) составляет 120 м, но впоследствии изменяется до 30 м для выравнивания с другими 5 диапазонами. На рис. 2 показаны временные окна работы различных поколений спутников Landsat.

В настоящее время Landsat 8 является предпочтительным спутником серии Landsat, поскольку он значительно более чувствителен, имеет дополнительные спектральные каналы и очень хорошо откалиброван. Предыдущие серии снимков Landsat следует использовать только для архивного анализа, чтобы обнаружить изменения во времени или частоту событий (смотрите Наблюдение за водой из космоса, веб-картирование наводнений GeoScience Australia <http://www.ga.gov.au/scientific-topics/hazards/flood/wofs>)

Спутники Sentinel-2 будут становиться все более актуальными, поскольку данные становятся все более доступными по всему миру.

Краткий обзор данных Landsat



Рисунок 2: Поколения спутников Landsat

Источник: <https://directory.eoportal.org/web/eoportal/satellite-missions/l/landsat-7>

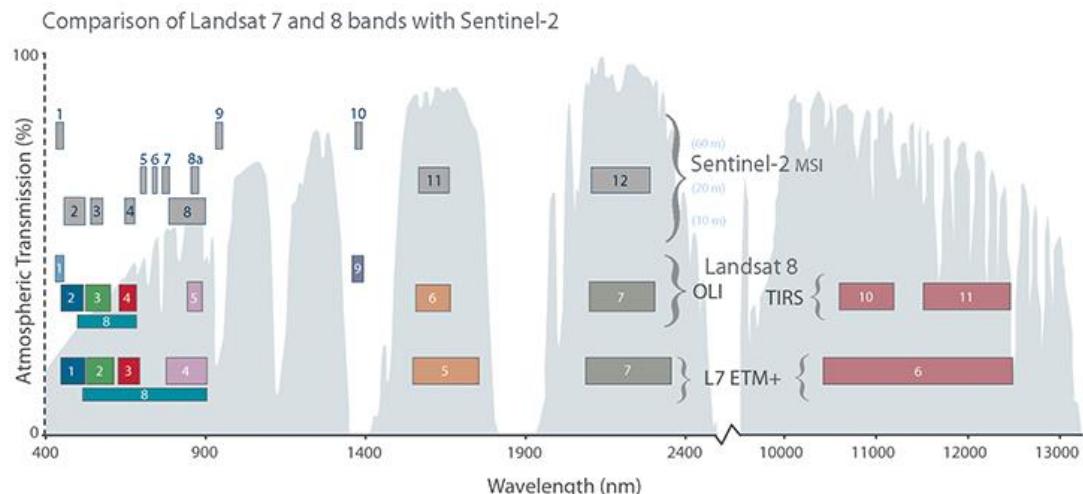


Рисунок 3: Спектральные каналы: Поколения спутников Landsat и Sentinel-2

Источник: <http://landsat.gsfc.nasa.gov/sentinel-2a-launches-our-compliments-our-complements/>

Спутниковые изображения Landsat записываются и сохраняются в виде рядов и колонках на основе конструкции, известной как Всемирная справочная система (WRS). Структурирование и сохранение изображений с помощью WRS таким образом позволяет запрашивать информацию о любом месте мира, указав номинальный центр сцены, обозначенный колонками и номерами строк (см. рис. 4).

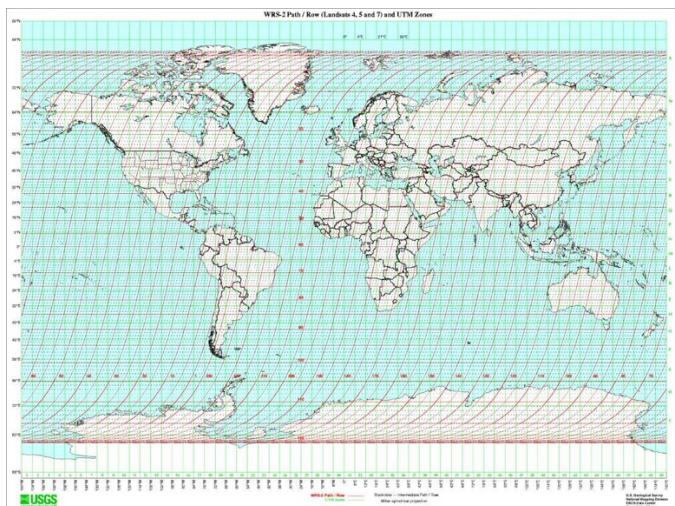


Рисунок 4: Всемирная справочная система

Источник: <http://landsat.gsfc.nasa.gov/the-worldwide-reference-system/>

Landsat 1-3 WRS-1 присваивает последовательные номера колонок с востока на запад 251 номинальной орбитальной траектории спутника, начиная с номера 001 для первой траектории, которая пересекает экватор на 65,48 градусах западной долготы. Конкретная орбитальная траектория может меняться из-за дрейфа и других факторов; таким образом, линия колонки является лишь приблизительной. Орбита периодически корректируется после того, как произошла определенная величина дрейфа, чтобы вернуть спутник на орбиту, которая почти совпадает с начальной орбитой. На рисунке 5 показан пример района Мельбурна, Австралия, отражающий сетку.

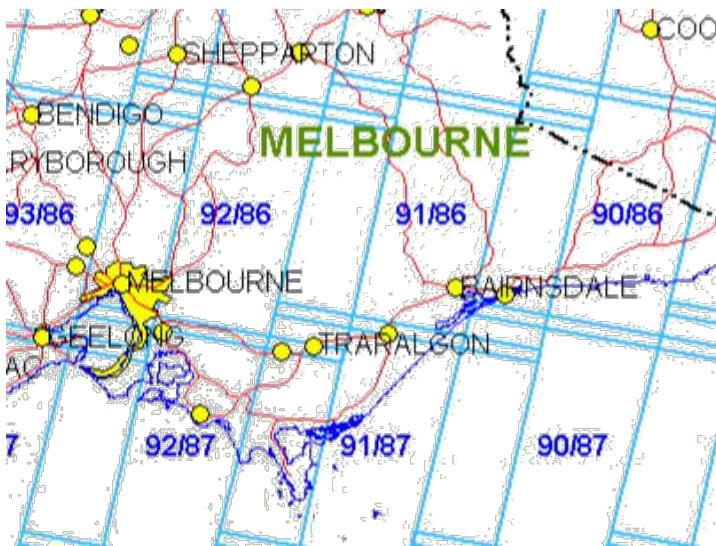


Рисунок 5: Спутниковые сетки колонок и рядов для выбранного района Мельбурна, Австралия.

Landsat 8 и Landsat 7 следуют за WRS-2, как и Landsat 5 и Landsat 4, тогда как Landsat 1, Landsat 2 и Landsat 3 последовали за WRS-1.

Landsat 4, 5 и 7 WRS-2 является расширением глобальных Landsat 1–3 WRS-1 и аналогичным образом использует упорядоченную систему колонок и рядов. Однако существуют серьезные различия в повторяющихся циклах, покрытии, схемах захвата и обозначениях колонки/ряда из-за больших различий в орбитах Landsats 4 и 5 по сравнению со Landsats 1–3.

После предварительной обработки спутниковых данных в готовые к анализу данные эта файловая система удаляется и заменяется каждым пикселим каждого изображения, помещаемым в

пространственно-временную базу данных, из которой можно выбрать любой фрагмент данных во времени и пространстве.

Спутники GEO расположены над экватором на высоте около 36 000 км, где они могут оставаться точно над одним и тем же местом на Земле. Первоначально предназначенные для использования в метеорологии, они становятся настолько сложными, что датчики также могут быть полезны и для других целей. Текущее самое высокое пространственное разрешение составляет 500 м для снимков, снятых с 10-минутными интервалами со спутника HIMAWARI-8, первого из серии подобных спутников (типа GOES-R), которые будут охватывать большую часть мира. Himawari-8 расположен над Индонезией и покрывает половину земного шара, хотя при большем отклонении от надира (от вертикали) спутник создает все более искаженные пиксели.

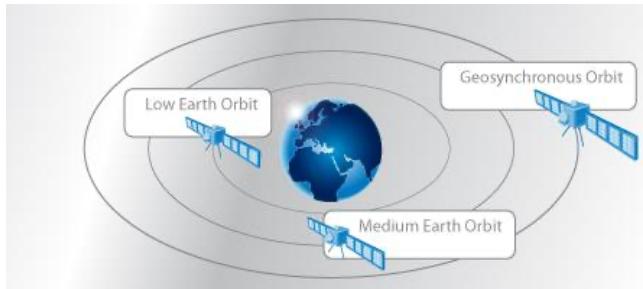


Рисунок 6: Визуальное представление спутниковых орбит

Источник: <http://www.idirect.net/Company/Resource-Center/~/media/Files/Corporate/iDirect%20Satellite%20Basics.pdf>

Спутники МЕО (средняя околоземная орбита, примерно 20 000 км) в основном ориентированы на навигацию, связь и геодезические/космические науки и имеют время обращения от 12 до 2 часов. Для получения дополнительной информации посетите веб-сайт NASA (Национальное управление по аeronавтике и исследованию космического пространства): <http://earthobservatory.nasa.gov/Features/OrbitsCatalog/>

2.3 Активные и пассивные датчики

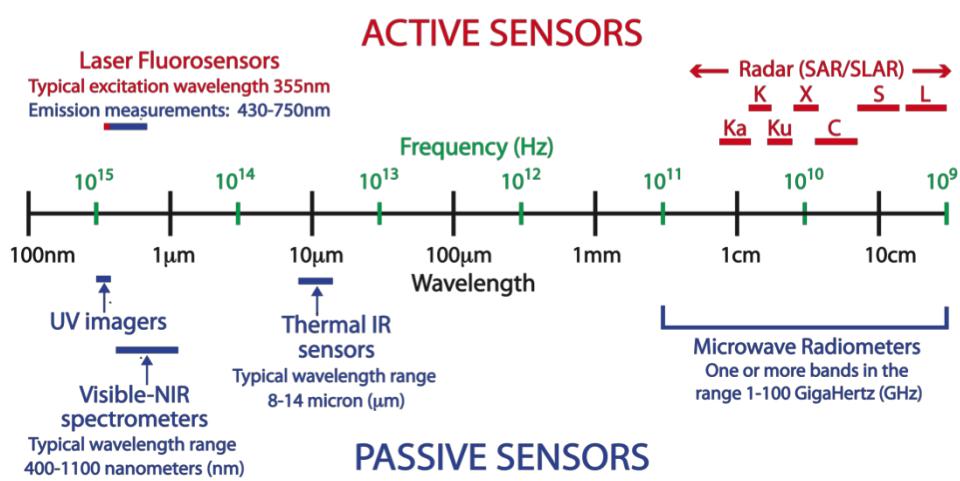


Рисунок 7: Характеристики активных/пассивных датчиков

Источник: http://lms.seos-project.eu/learning_modules/marinepollution/marinepollution-c01-s02-p01.html

Длины волн и частоты электромагнитных волн указаны в логарифмическом масштабе. Пассивные приборы обнаруживают сигналы, которые доступны естественным образом, например, отраженное

солнечное излучение или тепловое излучение, испускаемое целью. Активные приборы имеют собственный источник излучения для подсветки цели и обнаружения отраженного сигнала.

Также важно различать область электромагнитных длин волн, которую регистрируют эти датчики, и измеряют ли они отраженную солнечную энергию, излучения Земли (пассивные датчики наблюдения Земли) или сигнал, который они излучают, а затем записывают, когда он отражается Землей (активные датчики).

Для пассивных датчиков измеряемые длины волн отличаются от тех, которые может видеть человеческий глаз, например, от синего до зеленого, желтого, оранжевого и красного (длины волн видимого диапазона VIS Wavelengths); длины волн NIR (ближнего инфракрасного диапазона), SWIR (коротковолновые инфракрасные волны), MIR (средние инфракрасные волны в исследовательском режиме в настоящее время), TIR (тепловые инфракрасные волны), измеряющие температурное излучение до микроволнового диапазона (или радара). Датчики VIS, NIR, SWIR, MIR работают в основном в дневное время (за некоторыми исключениями, такими как использование ночных огней для обнаружения морских (судовых и энергетических платформ) или наземных источников искусственного света. Датчики TIR и пассивные микроволновые датчики также могут обнаруживать выбросы в ночное время. Пассивные микроволновые датчики могут использоваться для оценки влажности почвы на сушке или определения солености океана.

Отдельной недавней разработкой является обнаружение гравитационных аномалий с помощью спутников. Эти аномалии с течением времени могут, например, свидетельствовать об истощении или пополнении крупных подземных резервуаров подземных вод.

Активные датчики либо излучают свет с помощью лазеров в области VIS и NIR (LiDAR: технология обнаружения света и дальности действия), либо излучают микроволновую энергию с помощью радара с синтезированной апертурой (SAR). Эти системы могут работать днем и ночью, так как имеют собственный источник электромагнитной энергии. LiDAR используется для точной оценки высоты и определения высоты растительного покрова или глубины дна на мелководье (батиметрия). Исследования и разработки в области LiDAR сосредоточены на способности различать аспекты высоты растительного покрова или водной толщи.

SAR используется для отображения сигнала обратного рассеяния радара, по которому можно обнаружить и измерить толщину земного покрова, силу наводнения, определить сельскохозяйственные культуры, леса и т. д., а также незаконные и незарегистрированные суда. Интерферометрия SAR может обнаруживать миллиметровые сдвиги, такие как сдвиги в районах добычи ресурсов (проседание) или регионах с вулканической активностью или землетрясениями.

2.4 Применение алгоритмов к НЗ

НЗ обычно используются для определения пространственного распределения переменной (что и где?), а также для определения характера переменной. Например, цвет, глубина, потоки, энергия, концентрация, плотность, масса и т. д. Добавляя временной аспект повторного охвата, можно ответить на вопрос о состоянии и тенденции. Следовательно, эти измеренные отражения или выбросы необходимо преобразовывать в значимую информацию, например, для официальной статистики, посредством применения алгоритмов. В общем случае эти алгоритмы представляют собой математические уравнения, статистические или биофизические, геофизические, химико-физические методы или любую комбинацию этих подходов.

Характер этих алгоритмов имеет существенные последствия для их использования при подготовке официальной статистики. Можно выделить пять широких категорий с множеством возможных перестановок; первые три являются пиксельными методами, четвёртый использует эту информацию, а также пространственную и контекстную информацию. В пятом методе описаны гибридные подходы.

1. **Эмпирические методы:** когда устанавливается статистическая взаимосвязь между используемыми спектральными каналами или частотами и измеряемой переменной (на основе поля) без обязательной причинно-следственной связи. Этот метод наименее подходит для автоматизации на больших площадях и в различных условиях (угол солнца, время года, широта, наклон и экспозиция местности, атмосферные условия), если только он не сопровождается значительными текущими полевыми измерениями, предпочтительно при каждом прохождении спутника.
2. **Полуэмпирические методы:** когда устанавливается причинно-следственная связь между используемыми спектральными каналами или частотами и оцениваемой переменной. Этот метод менее склонен к получению ложных результатов, хотя результаты могут иметь высокий риск ошибки за пределами полевого диапазона. Этот метод имеет среднюю пригодность для автоматизации на больших площадях и предъявляет менее строгие требования к частому отбору проб в полевых условиях.
3. **Методы инверсии, основанные на физике:** также известные как полуаналитические методы инверсии. Все требуемые переменные оцениваются в одной спектральной инверсии одновременно. Этот метод обеспечивает согласованность результатов, основанную на физике, и наиболее подходит для автоматизации на больших площадях при условии, что модель инверсии правильно параметризована для требуемых переменных. Эти методы также называют биофизическими, биогеохимическими или геофизическими методами инверсии.
4. **Объектно-ориентированный метод анализа изображения (OBIA):** этот метод объединяет пространственную информацию, информацию о структуре, текстуре и спектре (например, посредством начальной сегментации изображения) с контекстной информацией, контролируемой человеком-оператором. Этот метод подходит для автоматизации естественных или искусственных систем, для которых он был разработан и все чаще используется.
5. **Методы искусственного интеллекта и машинного обучения:** это интеллектуальные многомерные статистические (часто нелинейные) модели для выявления очень сложных взаимосвязей в данных. В зависимости от того, как обучаются эти методы (с использованием полевых измерений или моделей, основанных на физике), они пересекают сферы эмпирических, полуэмпирических и физических методов инверсии на основе OBIA.

Методологические подходы к работе с данными НЗ более подробно обсуждаются в Главе 3.

Когда физические процессы хорошо изучены, мы можем разработать физические инверсионные модели. Часто будут использоваться префиксы, такие как биофизический или геофизический и т. д., указывающие на природу измеряемой переменной. Иногда данные НЗ могут обеспечить полезные входные данные для таких физических моделей, и это идеальная ситуация. Инвертирование физических моделей – это только один из способов их использования. Например, фракционный покров или индекс площади листьев (LAI) можно использовать в биофизической модели с опережением — для прогнозирования, например, валовой первичной продуктивности или эвапотранспирации. Другой пример – качество воды; некоторые аспекты которого могут быть определены из космоса с использованием моделей инверсии, основанных на физике. Преимущество этих инверсионных моделей состоит в том, что после правильной параметризации они больше не нуждаются в локальных данных и, таким образом, являются главными кандидатами для автоматизации обработки больших объемов данных НЗ. Поэтому, если это возможно, рекомендуются прямые и обратные подходы, основанные на физике, поскольку они уменьшают потребность в данных в месте нахождения и увеличивают возможность автоматизации.

Однако часто НЗ не могут дать прямых измерений входных данных, и именно тогда требуются полуэмпирические методы, которые находят корреляции между НЗ-производной переменной и какой-то реальной биофизической переменной. Когда физические процессы недостаточно хорошо изучены (например, экология), мы не можем использовать физические модели и прибегать к эмпирическим, полуэмпирическим или OBIA, или моделям машинного обучения, которые опираются на статистику для определения отношений, а не на физику. В этом случае эмпирические связи необходимо постоянно проверять (полевые работы или моделирование), чтобы убедиться, что они не меняются во времени и пространстве.

2.5 Способы комбинирования данных: слияние данных, слияние модели данных и ассилияция моделей-данных

Все шире используется объединение информации, содержащейся в двух или более датчиках НЗ, в целях устранения недостатков некоторых источников НЗ. Это называется слиянием данных или смешиванием. Примером такого подхода является использование наземных отражателей в VIS и NIR в сочетании с изображениями TIR или использование датчиков с высокой частотой повторения и низким пространственным разрешением с низкой частотой и высоким пространственным разрешением для компенсации недостатков друг друга.

Также возможно объединить информацию НЗ с базовой моделью изучаемого явления. Этот подход называется слиянием модели данных.

Когда данные НЗ и/или полевых наблюдений используются для повышения эффективности модели, а результаты моделирования затем используются для повышения достоверности или точности данных полевых наблюдений или продуктов НЗ или наоборот для поддержания согласованности между системными переменными, понимание неопределенности моделей и наблюдений, а также повышение информационного значения генерируемых выходов называется ассилияцией моделей-данных.

2.6 Актуальные примеры применения данных НЗ

Различные варианты применения геопространственной информации, полученной в ходе НЗ, существуют во многих областях, включая, помимо прочего, следующие сферы: биобезопасность, биоразнообразие, лесное хозяйство, сельское хозяйство, влажность почвы, водный баланс водосборных бассейнов, обмен энергией в атмосфере земли, цифровые модели рельефа, полезные ископаемые, управление природопользованием на шахтах, урбанизация, инфраструктура, товарные запасы, загрязнение окружающей среды, количество и качество воды, водоросли и коралловые рифы, батиметрия мелководья, продуктивность прибрежных и океанских вод, цветение водорослей, состояние моря, течения, зоны поднятия глубинных вод на поверхность и опускания поверхностных вод, качество атмосферного воздуха, управление чрезвычайными ситуациями и смягчение их последствий, управление рисками бедствий, экологический учет (т.е. мониторинг изменений земного покрова). Многие другие области применения будут разработаны по мере того, как датчики станут более сложными, а данные НЗ станут более доступными (см. раздел 2.8 о Готовых к анализу данных – ARD).

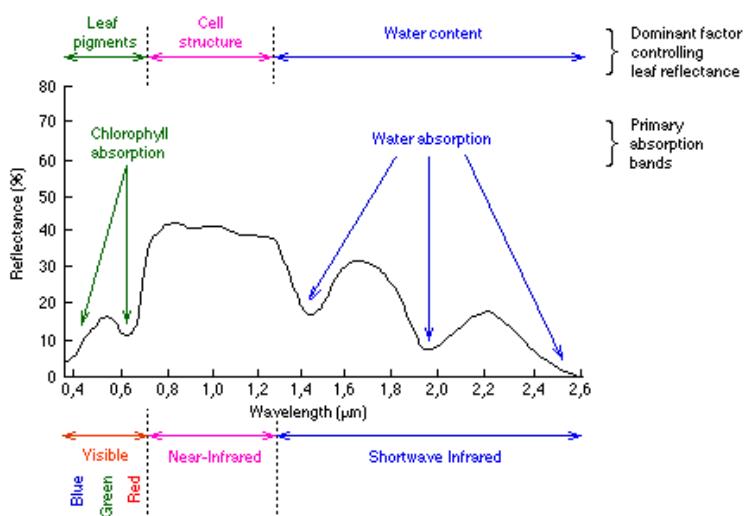


Рисунок 8: Основы дистанционного зондирования: характеристики длины волны и отражения

Источник: http://www.geog.ucsb.edu/~jeff/115a/remote_sensing/remotesensing.html

2.7 CEOS и GEO: международные организации по НЗ

GEO (Группа по наблюдению за Землей) и CEOS (Комитет по спутникам наблюдения Земли) являются двумя ключевыми игроками, связанными с НЗ; эти две глобальные организации тесно сотрудничают по различным вопросам, связанным с данными НЗ, а CEOS часто называют «космическим» подразделением GEO. Повестка дня на период до 2030 года по реализации и мониторингу ЦУР ООН – это прекрасная возможность отразить сотрудничество между двумя организациями, которые уже работают над инициативой GEO под названием EO4SDG, посвященной ЦУР ООН.

Космические агентства, обеспечивающие НЗ (через CEOS), и GEO признают, что им еще предстоит полностью устранить основные препятствия, с которыми сталкиваются потенциальные пользователи информации, полученной с помощью НЗ. Преобразование сигнала, измеренного в пространстве, в значимую переменную требует, чтобы этот сигнал прошел несколько этапов предварительной обработки, пока не будет рассчитан сигнал, из которого (в случае VIS, NIR или TIR) удаляются эффекты воздействия сенсорного шума, атмосферы, угла наклона солнца, угла рельефа, облаков и теней от облаков, затенения от конструкций с навесами или, в случае воды, отражений солнца и бликов неба.

CEOS был создан в 1984 году с первоначальной функцией координации и гармонизации НЗ, чтобы облегчить пользователям доступ к данным и их использование. Первоначально CEOS сосредоточился на функциональной совместимости, общих форматах данных, взаимной калибровке инструментов, а также на общей проверке и взаимном сравнении продуктов. Однако со временем обстоятельства, связанные со сбором и использованием космических данных НЗ, изменились. За последние три десятилетия CEOS внес значительный вклад в развитие усилий сообщества по наблюдению за Землей из космоса. Агентства CEOS (насчитывающие 31 участника) общаются, сотрудничают и обмениваются информацией о деятельности НЗ, стимулируя полезные партнерства, такие как Комплексная стратегия глобальных наблюдений (IGOS). Полный список членов CEOS доступен по ссылке: <http://ceos.org/about-ceos/agencies/>.

CEOS сыграл важную роль в создании и постоянном развитии GEO и GEOSS (Глобальная система наблюдения Земли). Агентства CEOS работают вместе, чтобы запустить межведомственные совместные миссии, и такие совместные усилия принесли большую пользу пользователям по всему миру. CEOS также предоставляет установленные средства связи с внешними организациями, что позволяет CEOS понимать потребности и требования этих организаций в области данных НЗ и действовать в соответствии с ними.

Основанная в 2005 году, GEO представляет собой добровольное партнерство правительств и организаций, которое видит «будущее, в котором решения и действия на благо человечества основываются на скоординированных, всесторонних и устойчивых наблюдениях Земли и информации». Более подробная информация доступна по ссылке: <https://www.earthobservations.org/vision.php>.

GEO выступила с инициативой («EO4SDGs» - Наблюдения Земли для достижения целей в области устойчивого развития на период до 2030 г.) для организации и реализации потенциала источников данных НЗ для продвижения Повестки дня на период до 2030 г. и обеспечения выгод для общества за счет достижения ЦУР. Инициатива EO4SDGS целенаправленно работает над расширением текущих партнерских отношений GEO, укреплением ее прочных отношений с ООН и содействует консолидированному участию отдельных стран-членов и организаций-участников. CEOS активно поддерживает эту инициативу GEO, а также создал специальную группу по ЦУР (<http://ceos.org/ourwork/ad-hoc-teams/sustainable-development-goals/>) для лучшей координации деятельности агентств CEOS по ЦУР.

GEO тесно сотрудничает с ключевыми структурами ООН, включая Статистический отдел ООН, Комитет экспертов ООН по управлению глобальной геопространственной информацией (UN GGIM) и Сетью ООН по выработке решений в области устойчивого развития (UN SDSN).

Дополнительные потенциальные партнеры включают банки развития, неправительственные организации (НПО) и международные организации. Взаимодействие и партнерские отношения с этими организациями помогают создавать процессы, механизмы и человеческий потенциал по включению данных НЗ в национальные планы развития и их интеграции с национальными статистическими счетами для улучшения измерения, мониторинга и достижения ЦУР.

Одно из направлений деятельности EO4SDG заключается в наращивании потенциала и взаимодействия; оказании поддержки странам в реализации всех соответствующих мер для надлежащего выполнения Повестки дня на период до 2030 года. Опираясь на деятельность по наращиванию потенциала в рамках GEO, эта активность координирует и поощряет усилия по наращиванию потенциала на соответствующих уровнях в отношении эффективных способов передачи методов, обеспечения доступа к данным и устойчивого использования НЗ в контексте ЦУР. Действия здесь основаны на усилиях GEO по определению потребностей пользователей и их поддержке. С учетом того, что в основе ЦУР лежат статистические данные, эта деятельность включает взаимодействие со статистическим сообществом ЦУР по вопросам НЗ, а также укрепление потенциала в рамках GEO и сообщества НЗ в отношении статистических принципов и практики ЦУР.

Таким образом, GEO и CEOS будут полностью вовлечены в процесс реализации ЦУР ООН, особенно при демонстрации того, где данные НЗ могут быть полезны для мониторинга некоторых ключевых показателей или для помощи в достижении целей/задач.

Некоторые проекты или инициативы уже имеют решающее значение для мониторинга НЗ, и их непрерывность поможет лучше связать сферы НЗ и официальной статистики. Например, см. раздел 2.14.2 пример GFOI (Глобальная инициатива по наблюдению за лесами), инициативы GEO.

Национальным статистическим управлением в странах, являющихся членами GEO, рекомендуется установить рабочие отношения со своими национальными космическими/метеорологическими/научными агентствами в области использования данных НЗ для целей официальной статистики.

2.8 Введение в данные НЗ на основе Готовых к анализу данных (ARD)

Многим пользователям спутниковых данных, особенно в развивающихся странах, не хватает опыта, инфраструктуры и пропускной способности интернета для эффективного доступа, подготовки, обработки и использования растущего объема исходных космических данных для принятия решений на местном, региональном и национальном уровнях.² Там, где такой опыт действительно существует, подготовка данных часто ограничивается небольшим числом специалистов, что ограничивает число конечных пользователей данных НЗ и/или использование разнообразных методов, что затрудняет надлежащее сравнение наблюдений и производных продуктов. Это приводит к часто несогласованным, а иногда и противоречивым результатам. Это серьезное препятствие для успешного использования космических данных и успеха крупных глобальных и региональных инициатив, поддерживаемых CEOS³ и GEO.

Поэтому полезно ввести понятие «Готовые к анализу данные» (ARD). Данные НЗ необходимо скорректировать с учетом многих эффектов, вызванных расстоянием до измеряемого объекта, и всех эффектов передачи, когда сигнал проходит от поверхности Земли и измеряется в космосе. Некоторые из этих эффектов связаны с атмосферой, солнечным углом, углом обзора от датчика до цели и т. д. Предварительная обработка данных для корректировки этих эффектов создает данные НЗ, как если бы датчик собирал данные в пределах нескольких метров от поверхности объекта.

Эта предварительная обработка данных НЗ может быть автоматизирована, что снижает нагрузку на специалистов по геопространственным данным, которые сосредоточены на извлечении ценной

² Полезная страница НАСА для получения информации о том, как использовать данные: <https://earthdata.nasa.gov/earth-observation-data/tools>

³ От проекта дискуссионного документа CEOS по анализу готовых данных (Управление по разработке систем CEOS 2016 г.)

информации из этих данных. Более того, если ARD хорошо определены и предпочтительно проверены экспертами, пользователь данных может быть уверен, что работает с данными, которые имеют приемлемо высокий уровень достоверности и точности.

CEOS предлагает рассмотреть следующее промежуточное определение:

«Готовые к анализу данные (ARD)» — это спутниковые данные, которые были обработаны в соответствии с минимальным набором требований и организованы в форме, позволяющей немедленно добавить ценность и провести анализ без дополнительных усилий со стороны пользователя».

Как только первичные данные обрабатываются до состояния ARD, становится более доступным и полезным правильное использование многочисленных технологий и инициатив, лежащих в основе «больших данных», которые позволяют проводить анализ и извлечение данных в реальном времени для повышения качества научных исследований. После достижения состояния ARD алгоритмы и методы, обсуждавшиеся ранее, могут быть применены и, что важно, с большей вероятностью будут сопоставимы для разных групп пользователей и конфигураций датчиков. Для НСУ в рамках получения показателей для статистических целей существует требование по доведению предварительно обработанных данных НЗ до уровня «Продукт», где они будут классифицированы по более значимым переменным, таким как классификация земного покрова. Это по-прежнему требует дополнительных знаний, которые могут быть вне или внутри НСУ. Для получения дополнительной информации об ARD см. раздел 2.10, а также Lewis и др., 2016 г.

2.9 Разнообразие источников данных НЗ (в местах нахождения, в воздухе)

Для работы с данными НЗ НСУ должны знать, какие источники можно использовать для мониторинга соответствующих переменных для официальной статистики. Помимо наиболее известных изображений MODIS и Landsat, существует множество источников доступных данных спутниковых изображений, бесплатных или коммерческих, с различных спутников, собирающих различные измерения о земле и земном покрове, воде и атмосфере. MODIS и Landsat являются наиболее используемыми источниками спутниковых данных из-за их раннего запуска (представляя относительно долгосрочные архивы) и долговечности, что приводит к значительному количеству рецензируемых публикаций, пространственному разрешению в сочетании с политикой открытых и бесплатных данных.

Однако два спутника MODIS, изначально предназначенные для использования в качестве исследовательских инструментов, будут заменены не идентичными системами, а системой под названием SUOMI-VIRRS, за которой последуют датчики NPOESS, объединяющие метеорологические и климатические наблюдения, а также наблюдения за сушей и океаном. Потенциал Landsat для мультиспектральных данных с разрешением 30 м больше не представляет максимально возможное пространственное разрешение данных в свободном доступе и больше не считается передовым в области разработки датчиков, хотя спутник Landsat 8 является значительным улучшением по сравнению с Landsat 5 и 7.

Ниже приведено руководство по выбору лучших спектральных каналов для использования Landsat 8.

Канал	Длина волны	Целесообразно для картирования
Канал 1 – прибрежная атмосфера со взвешенными частицами	0,43 – 0,45	Исследования мелководья, тонких частиц пыли
Канал 2 – синий	0,45 – 0,51	Батиметрическое картирование, отличающее почву от растительности и лиственных деревьев от хвойных
Канал 3 – зеленый	0,53 – 0,59	Акцентирует внимание на пике растительности, что полезно для оценки энергии прорастания растений. Общее содержание взвешенных частиц в водоемах

Канал 4 – красный	0,64 – 0,67	Различает спектральные наклоны растительности; также измеряет первичный фотосинтетический пигмент растений (наземных и водных): хлорофилл а
Канал 5 – ближний инфракрасный (NIR)	0,85 – 0,88	Подчеркивает содержание биомассы и береговую линию
Канал 6 – коротковолновый инфракрасный (SWIR) 1	1,57 – 1,65	Различает влажность почвы и растительности; проникает сквозь тонкие облака
Канал 7 – коротковолновый инфракрасный (SWIR) 2	2,11 – 2,29	Улучшенное содержание влаги в почве и растительности, а также проникновение сквозь тонкие облака
Канал 8 – панхроматический	0,50 – 0,68	Разрешение 15 метров, более четкое изображение
Канал 9 – Cirrus	1,36 – 1,38	Улучшенное обнаружение заслонения слоисто-перистыми облаками
Канал 10 – термоинфракрасный датчик TIRS 1	10.60 – 11.19	Разрешение 100 метров, тепловое картирование и оценка влажности почвы
Канал 11 – термоинфракрасный датчик TIRS 2	11.5 – 12.51	Разрешение 100 метров, улучшенное тепловое картирование и оценка влажности почвы

Таблица 1: На Landsat 8 установлены 2 инструмента: Operational Land Imager (OLI) – оперативный картограф Земли и Thermal Infrared Sensor (TIRS) – Тепловой ИК-сенсор

Источник: http://landsat.usgs.gov/best_spectral_bands_to_use.php плюс авторские дополнения (Dekker, A., Held, A. и Kerblat, F.)

Система спутников Copernicus Sentinel Европейской комиссии (спутники Sentinel типов 1-6, которые в настоящее время финансируются до 2030 года и управляются Европейским космическим агентством и EBMETCAT) представляет собой современные операционные системы со значительно расширенными возможностями. Возьмем, к примеру, спутник Sentinel-2 с 10-метровым разрешением и (запуск Sentinel 2B производится с 2017 года) повторяющийся цикл посещения одного изображения каждые пять дней.

Сейчас мы находимся на ранних стадиях технологической революции; глобальная сенсорная революция, в которой наземные, воздушные и космические датчики, а также новые формы анализа данных откроют совершенно новые подходы к глобальному управлению нашей планетой. Эти новые геопространственные технологии могут изменить точность, частоту, прозрачность и прогресс в оценке многих показателей ЦУР или других форм официальной статистики при значительно сниженных затратах на переменную на квадратный километр. НСУ и лица, принимающие решения, должны будут учитывать эту непрерывную технологическую эволюцию, включая демократизацию использования спутников. В дальнейшем рекомендуется, чтобы НСУ повышали уровень подготовки своих сотрудников по работе с этими источниками данных и постоянно обновляли эти навыки по мере изменения технологий.

2.10 Использование источников ARD или НЗ для официальной статистики

Преимуществом данных НЗ является их большой пространственный и временной охват, а их доступность, надежность и точность значительно улучшились. С помощью спутниковых данных НЗ почти 100% земной поверхности неоднократно покрывается, что намного больше, чем когда-либо было бы достигнуто с использованием полевых или воздушных технологий.

Недостатком НЗ является то, что, в отличие от полевых измерений или локальных измерений, оценки НЗ получаются косвенно (хотя можно утверждать, что это делают все сенсорные технологии, включая человеческий глаз, который фактически является системой дистанционного зондирования с тремя длинами волн, также называемым спектральными каналами). В то время как результаты НЗ для потенциального оперативного использования во многих областях выглядят многообещающе, в других же областях НЗ еще не удалось устранить все необходимые пробелы в

информации, и возможности для дальнейшего развития многих оперативных областей применения пока еще существуют.

Данные НЗ, особенно когда они используются вместе с данными, полученными в месте нахождения или модельными данными, могут внести существенный вклад в управление ресурсами и их разработку. Однако не всегда легко составить четкое представление о границах между взаимодополняемостью и взаимозаменяемостью этих методов. В этом заключается риск для обычного практика и планировщика; легко быть ошеломленным всеми этими достижениями и обещаниями, в то же время чувствуя себя ограниченным институциональной неспособностью справиться с длительными периодами времени и значительными бюджетными ресурсами, необходимыми для локальных измерений. Чтобы принимать обоснованные решения, специалисты-практики должны быть осведомлены как о потенциальной ценности, так и об ограничениях применения технологий и продуктов НЗ посредством обоснованных рекомендаций и разработки практических инструментов, ориентированных на результат. Дополнительная информация по этой теме находится в главе 5 на рисунках 32 и 33.

Сбор, использование, анализ и хранение больших данных – это проблема, которую можно, по крайней мере частично, решить с помощью спутников из-за их высокоструктурированного характера и того, что они предназначены для этой конкретной цели.

НСУ в настоящее время используют традиционные источники данных (опросы, выборки и т. д.), которые по-прежнему актуальны и полезны, но данные НЗ могут помочь завершить или получить более точные, частые, пространственно- или временно-экстраполированные, либо временно-интерполированные измерения.

В 2015 году Глобальная рабочая группа ООН по большим данным провела опрос, в котором задалась вопросом о стратегическом видении НСУ и их практическом опыте работы с большими данными. Интересно отметить, что среди 93 ответивших стран статистические управления считают «более быструю статистику», «снижение нагрузки на респондентов» и «модернизацию процесса производства статистических данных» – основными преимуществами использования больших данных. Затем последовали еще две причины для использования больших данных: «новые продукты и услуги» и «снижение затрат».

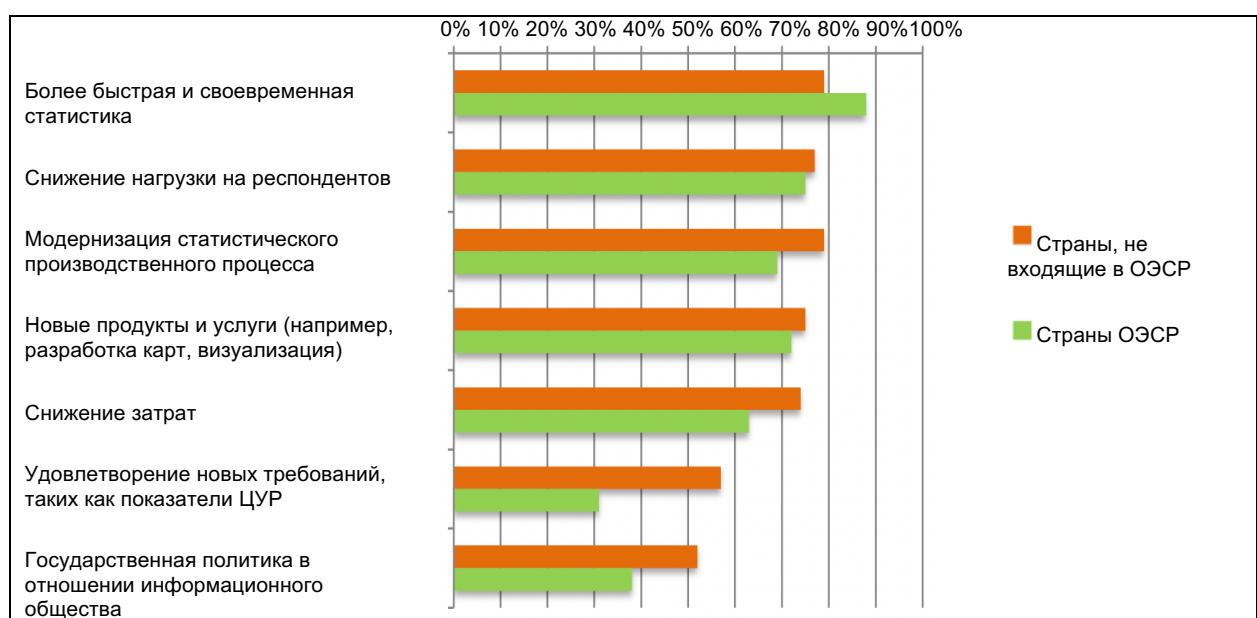


Рисунок 9: Основные преимущества больших данных для НСУ

Источник: Глобальный опрос, проанализированный в «Отчете Глобальной рабочей группы по большим данным для официальной статистики».

Информация, полученная с помощью НЗ, представляет собой растущую возможность дополнить традиционные источники (наземные или социально-экономические) при нехватке данных. Использование продуктов данных НЗ может сделать данные более плотными в пространственном или временном отношении (глобально), улучшить частоту или богатство данных и сэкономить деньги по сравнению с традиционными методами (методы опросов требуют больших затрат времени и средств). Данные НЗ могут изменить правила игры для мониторинга глобальных инициатив (таких как нынешние GFOI, GEOGLAM (GEO Глобальный мониторинг сельского хозяйства) и т. д.), и, следовательно, они могут быть полезными методами для мониторинга ЦУР ООН путем измерения и сбора некоторых соответствующих показателей. Если будет достигнута стадия автоматизированной системы обработки данных НЗ, обновления официальной статистики могут производиться чаще, например, ежемесячно.

Несмотря на потенциальные преимущества НЗ, это не всегда можно рассматривать, как наиболее подходящий источник для производства официальной статистики. Ниже приведены некоторые критерии, которые практикующие специалисты должны учитывать при оценке того, подходят ли данные НЗ для их целей. Дальнейший процесс анализа затрат и выгод описан в главе 5.

2.11 Достоверность информации и определения

Для эмпирических, полуэмпирических, ОВIA, AI и подходов машинного обучения существует три основных источника неопределенности в спутниковых оценках: (i) ошибки поиска, (ii) ошибки выборки и (iii) неадекватные наблюдения наземных ориентиров. В случае инверсионных моделей, основанных на физике (iv), параметризация модели может вносить неопределенности; однако эти методы лучше справляются с первыми тремя источниками неопределенности. Основанная на физике модель инверсии не обязательно нуждается в наземных наблюдениях для проверки достоверности после того, как она параметризуется репрезентативным образом на изменчивость целевых переменных. Это является дополнительным преимуществом в случае необходимости оценки переменных в отдаленных районах. Кроме того, основанные на физике методы инверсии, могут быть легко перенесены с одного датчика НЗ на другой без необходимости в новом комплексе наземных измерений.

Дополнительные источники неопределенности связаны с необходимостью калибровки модели, различными пространственными масштабами и необходимостью коррекции погрешности расчетных значений до их использования. Важная проблема при использовании оценок, основанных на дистанционном зондировании, заключается в том, как согласовать их с измерениями наземных ориентиров, поскольку и те, и другие могут быть наблюдениями самого разного характера. Таким образом, наблюдения из космоса необходимо анализировать, подтверждать и использовать в соответствии с их ограничениями.

Погрешности выборки и измерений могут быть результатом измерения переменной в неправильном месте (например, количество осадков, измеренное на нижней границе облаков, а не на поверхности земли) и косвенных оценок и погрешностей в измерительных датчиках, влекущих за собой ошибки в величине измерения (например, количество осадков). Эти ошибки будут зависеть от географических и атмосферных условий.

Тем не менее, базовые наземные измерения также имеют ошибки в отношении достоверности, репрезентативности, полноты, а в случае лабораторного анализа могут еще содержать ошибки подготовки проб, извлечения, хранения и измерения. Таким образом, как данные НЗ, так и данные о наземных ориентирах являются переменными с ошибками. По существу это модели погрешностей измерения, которые должны учитывать ошибки в переменных наземных ориентирах, часто рассматриваемых в качестве независимых переменных. Следует отметить, что в случае классификации сельскохозяйственных культур они не могут быть получены непосредственно на основе данных НЗ, а могут быть получены на основе данных о наземных ориентирах. Хотя погрешности измерений и будут иметь место, базовые справочные данные не требуют моделирования для превращения их в переменные, представляющие интерес для сельскохозяйственной статистики, что позволяет избежать ошибок моделирования.

Полезную публикацию о валидности информационных продуктов на основе НЗ (с акцентом на оценку гидрометеорологических переменных) можно найти в книге From Earth Observation for Water Resources: Current Use and Future Opportunities for the Water Sector (<https://openknowledge.worldbank.org/bitstream/handle/10986/22952/9781464804755.pdf>). Здесь дается краткое резюме с некоторой дополнительной информацией, относящейся к моделям инверсии, основанным на физике.

Несмотря на неопределенность, присущую локальным измерениям или измерениям наземных ориентиров, считается, что точность данных НЗ редко будет столь же высокой, как точность эквивалентных полевых измерений в этом месте. Тем не менее, не взирая на в целом более низкую точность привязки к конкретным участкам, продукты НЗ следует по-прежнему считать важным альтернативным источником данных, поскольку снимки НЗ могут обеспечивать информацию с большей пространственной протяженностью, пространственной плотностью и/или временной частотой, по сравнению с большинством данных, получаемых в ходе (точечных) наблюдений на местах. Кроме того, полученная с помощью НЗ информация может быть получена гораздо быстрее и является более экономически эффективной, чем большинство традиционных методов, основанных на обследованиях. После проверки эти данные можно также легко экстраполировать по крупным регионам и отдаленным районам с высокой эффективностью. Именно по этим причинам сочетание НЗ и полевых (наземные ориентиры) данных, как правило, обеспечивает наилучшие информационные результаты. Если есть возможность использовать модель (например, детерминированную), которая в достаточной степени описывает явление, представляющее интерес, то наиболее эффективным способом получения требуемой информации будет ассимиляция НЗ и данных о наземных ориентирах.

2.12 Когда данные НЗ считаются «действительными»? Аргумент целесообразности

Наиболее полезным определением в случае попытки установить достоверность информационного продукта на основе данных НЗ является концепция его соответствия назначению. Ученый, работающий с данными НЗ, как правило, пытается достичь высочайшего уровня точности, правильности и достоверности. Для статистиков, в контексте геопространственного анализа, пригодность информации, полученной в ходе НЗ, будет сосредоточена вокруг вопроса: есть ли преимущество по сравнению с существующими методами, не основанными на НЗ? Аспекты, которые могут быть рассмотрены, включают своевременность, пространственный и временной охват, оценку нескольких переменных из одного и того же набора данных, независимость респондентов, недопущение политических или других предубеждений и т. д.

Среди НСУ качество обычно понимается как «соответствие цели». Однако это предполагает оценку конечного результата с конкретной ссылкой на его предполагаемые цели или задачи. Таким образом, качество является многоаспектной концепцией, которая включает в себя не только точность статистических данных, но и другие составляющие, такие, как актуальность и интерпретируемость. За последнее десятилетие были предприняты значительные усилия для создания прочной основы для этих определений в статистике.

Например, Система оценки качества данных (DQF) Австралийского бюро статистики (ABS) 2009 года основана на Структуре обеспечения качества Статистического управления Канады (2002 г.) и Кодексе норм Европейской статистики (2005 г.). Помимо «Институциональной среды», были специально определены шесть параметров, способствующих обеспечению наилучшего качества данных: актуальность, своевременность, точность, согласованность, интерпретируемость и доступность. В следующей таблице представлены некоторые соображения о том, как эти измерения можно рассматривать с точки зрения НЗ, и добавлены два дополнительных определения из дисциплины НЗ: целостность и пригодность.

Таблица 2: Определения, данные статистиками и учеными по теме НЗ: сводная информация

	ABS DQF или сообщество НСУ	Ключевые аспекты (Как НСУ оценивают параметр?)	НЗ / сообщество дистанционного зондирования
Институциональная среда	«Институциональные и организационные факторы, которые могут оказывать существенное влияние на эффективность и авторитет агентства, производящего статистику».	<ul style="list-style-type: none"> • Беспрестрастность и объективность • Профессиональная независимость • Мандат для сбора данных • Достаточность ресурсов (для удовлетворения потребностей) • Обязательство по обеспечению качества • Конфиденциальность статистических данных 	<p>В целом данные НЗ беспристрастны, объективны и воспроизводимы. Обработка должна быть прозрачной и открытой для проверки.</p> <p>Космические агентства или коммерческие аэрокосмические компании предоставляют исходные данные «Верхней границы атмосферы». Эти же и другие агентства предварительно обрабатывают данные в формате, готовом для анализа. Впоследствии ARD необходимо преобразовать в соответствующую переменную либо средствами НСУ, либо для них.</p> <p>Критерии поиска: исчерпывающая информация о характеристиках датчика НЗ, его калибровке (предполетной, бортовой и вспомогательной); существует ли надежная рецензируемая литература? Очень важно иметь прозрачные алгоритмы и документацию, полные метаданные и документацию по продукту, продукты с открытым доступом или алгоритмы с открытым доступом.</p> <p>В данном случае концепция ARD имеет важное значение, поскольку она решает один аспект институционального потенциала: если эксперты мирового уровня проводят предварительную обработку данных НЗ, преобразуя их в кубы данных ARD НЗ, институциональный потенциал может сосредоточиться на применении правильного алгоритма, параметризации и проверке. Необходимо иметь потенциал геопространственной аналитики.</p>
Актуальность	«Насколько статистический продукт или публикация отвечает потребностям пользователей с точки зрения измеряемой	<ul style="list-style-type: none"> • Охват и среднее значение⁴ • Отчетный период⁵ • Географические детали • Основные результаты / элементы данных 	Отчетный период = срок службы записи, т. е. срок работы программы НЗ. Это может быть один спутниковый датчик или набор датчиков, таких как Landsat (несколько датчиков с 1984 года).

⁴ Цель или задача сбора информации, включая определение целевой группы населения, обсуждение вопроса о том, кого представляют данные, кто исключается из процесса и существуют ли какие-либо последствия или предубеждения, вызванные изоляцией конкретных людей, районов или групп

⁵ Период, за который были собраны данные, а также были ли какие-либо исключения в отношении периода сбора

	концепции и представленного населения».	<ul style="list-style-type: none"> • Классификация и статистические стандарты • Типы имеющихся оценок • Другие источники⁶ 	<p>Географическая детализация соответствует пространственному разрешению датчика НЗ, которое необходимо выбрать для предполагаемой цели.</p> <p>Временное разрешение также важно для применения НЗ и зависит от датчика.</p> <p>Основные результаты/элементы данных и типы оценок: см. обсуждение различных методов и алгоритмов в главе 3.</p> <p>Существует много методов классификации НЗ, но есть потребность в широко применяемых стандартах (например, GFOI и GEOGLAM). Потенциально НСУ могли бы разработать стандарты классификации для этого типа данных.</p> <p>Доступные типы оценок будут улучшены за счет увеличения: географического охвата (глобальный, региональный, по запросу) и варианты разрешения датчика; пространственные, спектральные, временные, радиометрические.</p> <p>Другие источники: вопрос, который необходимо рассмотреть, заключается в том, имеются ли минимальные или достаточные данные о наземных ориентирах.</p> <p>Смежным понятием в НЗ является целостность; биофизическая и научная целостность. Это касается способности соотносить и интерпретировать производные метрики НЗ с биофизическими процессами.</p>
Своевременность	«Задержка между базисным периодом (к которому относятся данные) и датой, на которую они были получены; задержка между объявленной датой и датой получения данных».	<ul style="list-style-type: none"> • Временная привязка⁷ • Частота обследований 	<p>С автоматизированными системами обработки, производящими 1) ARD и 2) должным образом проверенные алгоритмы, становится возможной доставка почти в реальном времени, т.е. в срок от нескольких дней до нескольких часов после приема данных НЗ.</p> <p>В сфере НЗ есть такое важное понятие как пригодность: это функция</p>

6

www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1520.0Main%20Features4May%202009?opendocument&tabname=Summary&prodno=1520.0&issue=May%202009&num=&view=

⁷ Задержка времени между базисным периодом и моментом фактического получения данных (включая задержку между объявлением датой выпуска и фактической датой выпуска)

			точности и временной привязки. Высокая точность уже после события гораздо менее полезна, чем более низкая точность в начале или до определенного биологического периода, например посев, цветение урожая или сбор урожая или, например, в случае коралловых рифов – их обесцвечивания.
Точность	<p>«Степень, в которой данные правильно описывают явление, для измерения которого они предназначены».</p> <p>«Должно оцениваться с точки зрения основных источников ошибок, которые потенциально могут привести к неточности».</p>	<ul style="list-style-type: none"> • Ошибка покрытия • Ошибка выборки • Ошибка неответа • Ошибка ответа • Другие источники⁸ 	<p>Что касается производных информационных продуктов НЗ, то в первую очередь необходимо установить их достоверность. Таким образом, важно установить причинно-следственную связь между измерением НЗ и данным явлением. Эмпирические алгоритмы могут подвергнуться риску получения ложных результатов.</p> <p>Данные НЗ имеют высокий пространственный и временной охват, но могут иметь пробелы из-за малого числа перепроверок и/или неблагоприятных погодных условий или других условий изображения.</p> <p>Для результатов классификации (продукты с категориальными переменными, например, земной покров, тип растительности, бентический тип) необходимо определить пользователей, производителей и общую точность. Обратите внимание, что каппа-статистика считается устаревшей.</p> <p>Для продуктов с непрерывными переменными (например, дробное покрытие, концентрация хлорофилла-а) используются многочисленные статистические данные. Общие статистические данные включают RMSE и ее варианты, смещение, коэффициенты R², r-Пирсона.</p> <p>Что касается RMSE и систематического отклонения, то результаты могут существенно варьироваться в зависимости от того, какое значение ставится на первое место; НЗ или наземные ориентиры. Коэффициенты R² и R затрудняют понимание, насколько хорошо данные совпадают (данные могут быть сильно</p>

			<p>линейно коррелированы и не располагаться «один-к-одному»).</p> <p>Важно оценивать продукты НЗ с помощью независимых данных. Вместе с тем следует отметить, что измерения наземных ориентиров также содержат факторы неопределенности.</p>
Согласованность	«Внутренняя согласованность сбора статистических данных, продуктов или публикаций, а также их сопоставимость с другими источниками информации в широких аналитических рамках и с течением времени».	Изменения в элементах данных Сравнение между элементами данных Сравнение с предыдущими выпусками Сравнение с другими доступными продуктами	<p>Наборы данных НЗ являются внутренне весьма непротиворечивыми, поскольку все они основаны на измерениях, производимых одним и тем же прибором. Тем не менее, непрерывно улучшается обработка от первичных данных верхней части атмосферы к ARD. Аналогичным образом, алгоритмы НЗ постоянно совершенствуются. Более широкое использование данных НЗ, базовых данных и данных моделирования повышает достоверность и точность, однако может усложнить сопоставление с данными, полученными с помощью предыдущей версии.</p> <p>Проблема непротиворечивости также возникает, когда используются многочисленные источники данных НЗ.</p>
Интерпретируемость	«Доступность информации, которая помогает понять данные». ⁹	Презентация информации Доступность информации о данных	Интерпретация информации, полученной с помощью НЗ, требует нескольких навыков в области обработки данных НЗ, а также в области геопространственной аналитики. Данные НЗ часто сводятся к таблице и графике, уничтожая часть интересной пространственной и временной контекстуальной информации. Ключевое значение имеют надлежащая документация и ведение учета. Требуется руководство по вопросу о сфере охвата и ограничениях использования данных НЗ.
Доступность	«Легкость доступа к данным для пользователей, включая легкость, с которой может быть установлено существование информации, а также пригодность формы или средства, с помощью которого можно получить	Доступность для общественности (степень доступности данных для общественности или уровень ограничений доступа)	Информационные продукты НЗ, как правило, являются весьма крупными и сложными, что затрудняет их доступность. Информационные продукты НЗ должны быть доступными, а это означает, что они должны предоставляться в формате данных, стандартизированном и легко доступном для различных видов программного обеспечения, особенно программного обеспечения открытого доступа. Желательно также, чтобы информационные продукты имели

⁹ Информация для облегчения толкования может включать используемые переменные, наличие метаданных, включая концепции, классификации и показатели точности.

	<p>доступ к информации».</p> <p>«Стоимость информации может также представлять собой один из аспектов доступности для некоторых пользователей».</p>	<p>Доступные информационные продукты¹⁰</p>	<p>открытый доступ (GEO). Полезно, если данные НЗ предоставляются в соответствии с открытыми геопространственными стандартами (OGS) и стандартами геопространственного программного обеспечения и данных с открытым исходным кодом (OSGeo).</p> <p>Еще одной стандартизацией является Библиотека абстракции геопространственных данных (GDAL), которая представляет собой компьютерную программную библиотеку для чтения и записи растровых и векторных геопространственных данных.</p> <p>Существует совместная рабочая группа по стандартам OGC и W3C, которая разрабатывает подход к публикации данных НЗ в качестве "увязанных данных" для облегчения доступа, используя модель для кубов данных, которая была первоначально разработана статистическим сообществом SDMX.</p> <p>Если ARD будут приняты всеми до преобразования этих данных ARD в значимый информационный продукт, интерпретируемость увеличится, поскольку будет легче создавать растущий объем знаний. Это также приведет к большей стандартизации продуктов.</p>
--	---	---	---

Поскольку использование НЗ НСУ является относительно новым, рекомендуется проводить пилотные и демонстрационные исследования, в которых изучаются эти определения и уточняется приведенная выше таблица. Некоторые пилотные исследования, проведенные членами Целевой группы ООН по спутниковым изображениям и геопространственным данным, описаны в главе 4.

В следующем разделе будет показано, что потенциал использования спутниковых данных НЗ намного больше, чем использование только временных рядов изображений MODIS или LANDSAT. Доступно больше источников, и будут представлены примеры того, какие датчики лучше всего подходят для получения наилучшего ожидаемого качества данных.

¹⁰ Имеющиеся конкретные продукты (например, публикации, электронные таблицы), форматы этих продуктов, их стоимость и имеющиеся в них элементы данных.

2.13 Доступные источники информации о данных наблюдения Земли

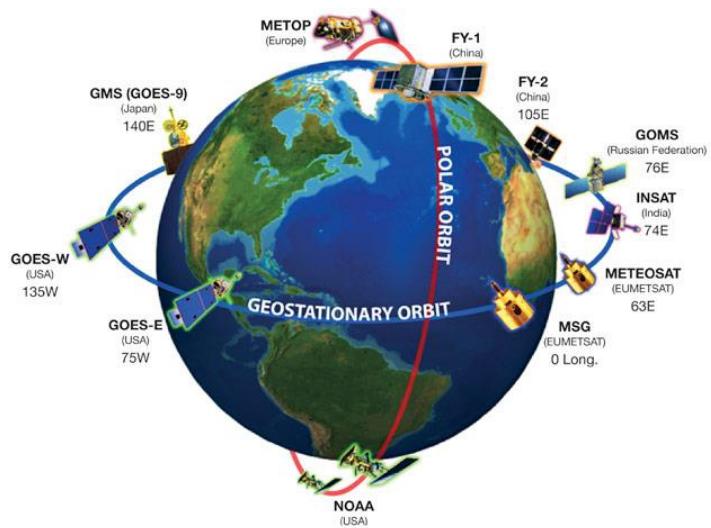


Рисунок 10: Операционные метеорологические спутники

Источник: Справочник CEOS EO: http://www.eohandbook.com/eohb2011/climate_satellites.html

Полный список спутников Земли прошлого, настоящего и будущего существует в онлайн базе данных OSCAR Всемирной метеорологической организации (WMO), которая содержит 645 записей (по состоянию на апрель 2016 г.). <https://www.wmo-sat.info/oscar/satellites>

Существуют фильтры для сортировки данных по:

1. Акронимам платформы датчиков наблюдения Земли (которая может содержать несколько бортовых аппаратур: один элемент полезной нагрузки – это конкретный датчик наблюдения за Землей),
2. Дата запуска и окончание срока службы;
3. Название программы космического агентства;
4. Список орбитальных характеристик и
5. Бортовая аппаратура.

В качестве примера можно привести спутник Sentinel-3 (очень важный и перспективный на следующие 15 лет для мониторинга суши и океана в видимом и ближнем инфракрасном диапазоне, а также для измерения температуры поверхности земли и океана с помощью наземного и океанического высотомера).

В системе есть две спутниковые платформы Sentinel-3 А и В; одна была запущена 16 февраля 2016 года, а другая – в 2017 году. После этого уже профинансированные аналогичные Sentinel-3 С и D придут на замену Sentinel-3 А и В. Sentinel E и F также запланированы, но смогут улучшить работу датчиков только на основе результатов предыдущих 4-х спутников. Sentinel-3 А и В имеют ожидаемый срок службы 7 лет. Для некоторых датчиков НЗ ожидаемый срок службы может быть превышен в 3-4 раза – космические агентства часто проявляют осторожность в этом отношении. Спутник летит по солнечно-синхронной орбите, что означает, что он будет проходить точно по той же полосе земли в то же самое солнечное время, на высоте 814,5 км. Статус спутника Sentinel 3 А – в работе, а Sentinel 3 В – запланирован. Бортовая аппаратура состоит из 7 датчиков: DORIS, GPS, LRR, MWR, OLCI, SLSTR и SRAL. Первые четыре датчика используются главным образом для точного позиционирования и оценки атмосферных условий, в то время как последние три являются фактическими датчиками наблюдения Земли с прикладным применением для обнаружения и мониторинга Земли; OLCI (спектрометр для измерения цвета океана и земли), SLSTR (радиометр

температуры поверхности моря и суши) и SRAL (радиолокационный высотомер с синтезированной апертурой).

В базе данных OSCAR Всемирной метеорологической организации, выбирая акроним датчика НЗ, пользователь переходит на страницу прибора, где представлен большой объем информации о датчике. Представлена, кроме того, информация о других спутниках, на которых также летает этот датчик; его вклад в космический потенциал и весьма информативная «предварительная оценка измерений», содержащая перечень измерений, которые обычно могут быть получены с помощью этого типа инструментов. Для OLCI эта таблица выглядит следующим образом¹¹.

Переменная	Релевантность для измерения этой переменной	Эксплуатационные ограничения	Уровень обработки
<u>Коэффициент диффузного отражения океана (DAC)</u>	1- первостепенная	Облака	Консолидированная методология
<u>Цветное растворенное органическое вещество (CDOM)</u>	1- первостепенная	Облака	Консолидированная методология
<u>Концентрация хлорофилла в океане</u>	1- первостепенная	Облака	Консолидированная методология
<u>Концентрация взвешенных отложений в океане</u>	1- первостепенная	Облака	Консолидированная методология
<u>Нормализованный разностный вегетационный индекс (NDVI)</u>	1- первостепенная	Облака	Консолидированная методология
<u>Разлив нефти</u>	2- очень высокая	Облака	Консолидированная методология
<u>Верхняя граница облаков</u>	3- высокая	Облака	Консолидированная методология
<u>Способность облаков отражать короткие волны</u>	3- высокая	Нет определенных ограничений	Консолидированная методология
<u>Аэрозольный вулканический пепел (общий столб)</u>	3- высокая	Облака	Консолидированная методология
<u>Морской ледяной покров</u>	3- высокая	Облака	Консолидированная методология
<u>Аэрозольная оптическая толщина</u>	3- высокая	Облака	Консолидированная методология
<u>Оптическая плотность облака</u>	3- высокая	Облако не слишком плотное	Консолидированная методология
<u>Доля поглощённой фотосинтетически активной радиации (FAPAR)</u>	3- высокая	Облака	Консолидированная методология
<u>Интегрированная масса водяного пара (IWV)</u>	3- высокая	Землепользование Облака	Консолидированная методология

¹¹ Источник: <https://wmo-sat.info/oscar/instruments/view/374>: Предварительная оценка результатов измерений.

Следующий список указывает, какие измерения обычно можно получить из этой категории приборов. Для просмотра полного анализа пробелов по переменным щелкните соответствующую переменную. Обратите внимание: таблица может быть отсортирована по заголовкам столбцов.

<u>Альbedo земной поверхности</u>	3- высокая	Облака	Консолидированная методология
<u>Коэффициент коротковолнового излучения в верхней части атмосферы</u>	4- умеренная	Нет определенных ограничений	Консолидированная методология
<u>Конденсированные частицы в атмосфере</u>	4- умеренная	Облака	Консолидированная методология
<u>Фотосинтетически активная радиация (PAR)</u>	4- умеренная	Облака	Консолидированная методология
<u>Облачный покров</u>	4- умеренная	Нет определенных ограничений	Консолидированная методология
<u>Типы аэрозоля</u>	4- умеренная	Облака	Сильно зависит от внешней информации
<u>Эффективный радиус распространения аэрозоля</u>	4- умеренная	Облака	Сильно зависит от модели
<u>Аэрозольный вулканический пепел</u>	4- умеренная	Облака	Консолидированная методология
<u>Плотность потока коротковолнового излучения на поверхности Земли</u>	5- низкая	Только общий столб. Облака	Сильно зависит от внешней информации
<u>Соотношение массы конденсированных частиц в атмосфере</u>	5- низкая	Облака	Консолидированная методология

Рисунок 11: Предварительная оценка результатов измерений

Данный список указывает, какие измерения обычно можно получить из этой категории приборов. Для просмотра полного анализа пробелов по переменным щелкните соответствующую переменную. Обратите внимание: таблица может быть отсортирована по заголовкам столбцов. Интересной особенностью этого сайта является страница анализа пробелов: <https://www.wmosat.info/oscar/gapanalyses>, где, если выбрать, например, «Поверхность земли», появится меню приложений в контексте поверхности земли, например, «тип почвы», которое затем составляет список инструментов, их релевантность для измерения запрошенной переменной (с цветовой кодировкой), а под этим исчерпывающим списком другой список, упорядоченный с точки зрения релевантности, уровня обработки и имитации операций.

Еще одну базу данных спутниковых миссий можно найти на сайте CEOS:
<http://database.eohandbook.com/database/missiontable.aspx>

Она организована немного по-другому: название миссии является начальным полем (эквивалент платформы датчиков наблюдения за Землей в базе данных Oscar). Затем поля заняты следующей информацией: ответственное космическое агентство или коммерческий оператор, статус, дата запуска, данные об окончании срока службы, затем приложения, за которыми следуют инструменты, которые выполняют измерения для этих приложений, а следом сведения об орбите и URL-адрес.

Привлекательной особенностью этой базы данных является возможность поиска по ключевым словам, например, «область применения»:

<http://database.eohandbook.com/measurements/overview.aspx>

Также имеется краткий информативный обзор измерений, разделенных на категории атмосферы, земли, океана, снега и льда, а также гравитационных и магнитных полей. В каждой категории справа виден небольшой график с временными шкалами, и, если он выбран, откроется график,

показывающий полные временные шкалы прошлых, настоящих и запланированных датчиков, что позволяет оценить непрерывность этого конкретного потока данных. При выборе, например, в разделе «Земля» категории «растительность» открывается новая страница со всеми параметрами растительности, которые можно измерить.

Например, если выбран параметр Measurements > Land > Vegetation > Chlorophyll Fluorescence from Vegetation on Land, будет представлен обзор приборов для наблюдения за Землей с кратким описанием предполагаемых сфер применения и указанием запланированных инструментов, предназначенных для оценки углерода путем измерения флуоресценции хлорофилла, индуцированной солнцем.

Ниже приведен список нескольких репозиториев спутниковых снимков, данные из которых можно скачать бесплатно.

- USGS Earth Explorer;
- ESA Sentinel Mission;
- NOAA CLASS;
- NASA Reverb;
- Earth Observation Link (EOLi);
- National Institute for Space Research (INPE), специально для Южной Америки и Африки; Bhuvan Indian Geo-Platform of ISRO;
- JAXAs Global ALOS 3D World;
- NOAA Data Access Viewer Discover Authoritative Datasets;
- VITO Vision Coarse Vegetation Data;
- NOAA Digital Coast Snorkel the Seashore;
- Global Land Cover Facility Derived Satellite Data
- DigitalGlobe Free Product Samples;
- BlackBridge/Geo-Airbus Tag Team Champions of Satellite Imagery;
- UNAVCO Research Data.

USGS Earth Explorer – одно из самых известных и популярных агентств по свободному доступу к данным спутника Landsat. Пользователи должны зарегистрироваться, чтобы получить доступ к хранилищу данных Геологической службы США, и эта регистрация бесплатна.

Для получения подробной информации о подходящих датчиках НЗ в прошлом, настоящем и будущем для водных ресурсов см. открытый отчет, написанный для Всемирного банка¹². В нем представлен обзор проблемных областей, связанных с управлением водными ресурсами, и обсуждаются требования к данным для решения этих проблем с акцентом на те переменные, которые могут быть получены или оценены с помощью НЗ. Помимо таких характеристик поверхности, как топография, оседание почвы и другие, НЗ может охватывать восемь ключевых гидрометеорологических переменных, имеющих отношение к использованию водных ресурсов; осадки, эвапотранспирация, влажность почвы, растительность, землепользование и земной покров; грунтовые воды, поверхностные воды, снег и лед, а также качество воды.

¹² García, Luis E., Diego J. Rodríguez, Marcus Wijnen, and Inge Pakulski, eds. Earth Observation for Water Resources Management: Current Use and Future Opportunities for the Water Sector. Washington, DC: World Bank Group. doi:10.1596/978-1-4648-0475-5. License: Creative Commons Attribution CC BY 3.0 IGO

2.14 Характеристики продукта данных

Типы данных, полученных в ходе НЗ (пример управления водными ресурсами)

Тип данных	Описание
Исходные данные	Измерения датчика, полученные непосредственно со спутника, отформатированные как «цифровые отсчеты».
Обработанные данные	a) Сигнал ТОА (верхняя граница атмосферы). Исходные данные, преобразованные в данные ТОА: преобразование в реальные единицы измерения, такие как излучение ($\text{Вт}\cdot\text{м}^{-2}\cdot\text{нм}^{-1}\cdot\text{ср}^{-1}$) или коэффициент отражения (%); калибровка сигнала. б) Поверхностный сигнал. Данные ТОА обрабатываются в соответствии с данными, эквивалентными поверхности: корректировки, применяемые для устранения эффектов, связанных с углом обзора с помощью атмосферных и солнечных датчиков; сшивание сцены; геолокация и повторная проекция.
Продукты данных	Преобразование обработанных данных в продукты, описывающие реальные (обычно биофизические) переменные, такие как концентрация хлорофилла, площадь листьев, интенсивность осадков, температура поверхности и масса влажности почвы.
Информационные продукты	Преобразование продуктов данных в важную для управления информацию для поддержки принятия решений, например, состояние эвтрофикации озера, риск затопления дельты реки и устойчивые коэффициенты орошения в сельскохозяйственных районах.

2.14.1 Ключевые характеристики продуктов данных

Пространственное разрешение (зависит от датчика)

Пространственное разрешение данных НЗ, иногда называемое пространственной частотой или разрешением изображения, относится к размеру изображения в пикселях. Пространственное разрешение данных определяет точность, с которой могут быть зафиксированы пространственные изменения наблюдаемого явления. Относительно большой размер пикселя захватит меньше мелкомасштабных пространственных вариаций, чем относительно небольшой размер пикселя. Таким образом, пространственное разрешение определяется как размер наименьшего отдельного компонента или точки (называемой пикселием), из которой состоит изображение. Например, если разрешение спутника указано как «5 метров», каждый пиксель изображения имеет размер 5 на 5 метров.

Другой аспект – это ширина полосы захвата; она эквивалентна количеству пикселей в направлении движения спутника. Например, если прибор имеет 1000 пикселей размером 5 м, то ширина его полосы захвата составляет 5 км. Если спутниковый датчик имеет 2000 пикселей размером 1 км, то ширина полосы захвата составляет 2000 км. Большая ширина полосы захвата связана с более высокой частотой повторного охвата для одного датчика на одном спутнике. При наличии нескольких спутников скорость покрытия увеличивается с каждым датчиком на каждом спутнике.

Когда пиксель охватывает два (или более) отдельных объекта земли (например, водоем и прилегающую растительность), пиксель фиксирует смесь сигналов от обоих объектов и иногда называется «микселям». Миксели делают интерпретацию изображения более сложной, и чем больше размер пикселя, тем больше микселей могут содержать данные. Смешанные пиксели описаны в Главе 3.

Временная частота (зависит от датчика)

Временная частота данных (или временное разрешение) относится к тому, как часто датчик производит наблюдения за данным местом. В случае спутников на полярной орбите частота связана с частотой прохождения и обычно измеряется в днях. С текущими низкоорбитальными группировками в некоторых случаях может быть достигнута субдневная пропускная способность. Частота получения изображений с геостационарных спутников значительно выше и измеряется в порядке от нескольких минут до нескольких часов, поскольку датчик находится в неподвижном положении над фиксированной точкой над экватором. Расстояние от геостационарных датчиков до земли (~ 36000 км) означает, что пространственное разрешение является низким (в настоящее время между 0.5 и 2 км). Наблюдения с относительно высокой частотой лучше отражают динамику быстро меняющихся процессов, чем наблюдения с относительно низкой временной частотой.

Для некоторых областей применения время наблюдения может быть важным либо для того, чтобы наблюдения проводились в одно и то же время каждый день, либо потому, что наблюдения должны происходить в определенное время дня.

Длина записи (зависит от датчика)

Длина записи относится к длине записи данных. Обычно это зависит от периода эксплуатации спутника (или серии подобных спутников) и, таким образом, определяется датами запуска и окончания миссии. Однако, часто спутниковые датчики получают данные намного позже запланированных сроков завершения миссии (например, Landsat-5).

Вообще говоря, чем больше длина записи, тем старше спутник и связанная с ним технология. Таким образом, обычно существует компромисс между длиной записи и другими атрибутами данных, такими как точность, пространственное разрешение и/или количество спектральных каналов. С другой стороны, долговечные сенсорные системы и их потоки данных часто создают значительные возможности для разработки оперативных методов извлечения информации, подкрепляемых соответствующей научной и прикладной литературой.

Требования к полевым данным или данным наземных ориентиров (зависит от продукта)

Некоторые продукты данных НЗ создаются исключительно на основе спутниковых наблюдений, некоторые – на основе комбинации спутниковых и полевых наблюдений. Последние предъявляют гораздо более высокие требования к входным данным и поэтому больше зависят от наличия подходящих полевых данных. Данные традиционных обследований, собранные НСУ, в некоторых случаях могут предложить подходящие данные для проверки на местах. Данные НЗ, полученные в результате комбинации спутниковых и полевых наблюдений, также, как правило, более сложны для получения и могут быть ограничены конкретными местоположениями и временем, которые представляют полевые данные.

- **Низкие** требования к полевым данным относятся к продуктам данных, которые не используют полевые данные или используют их только для начальной параметризации алгоритма или для целей проверки выходных данных алгоритма;
- **Средние** требования к полевым данным относятся к продуктам данных, которым требуются полевые данные для калибровки данных НЗ, или которые используют умеренное количество полевых данных для получения конечного продукта НЗ (например, когда данные речных датчиков объединяются со спутниковыми данными с целью получения информации о масштабах паводков для оценки их объемов);

- **Высокие** требования к полевым данным относятся к продуктам данных, которые включают несколько источников полевых данных (например, большинство продуктов данных об эвапотранспирации (ET) и влажности почвы), иногда в комплексных системах асимиляции данных.

Надежность источника продукта данных (зависит от продукта)

Надежность продукта данных относится к уверенности в поставке этого продукта в пространстве и во времени. Чем больше пространственный охват, тем чаще обновляется продукт; чем больше вариантов получения продукта, тем выше надежность продукта.

- **Низкая** надежность описывает продукт, созданный специально для определенного времени, региона, применения или созданный только одной организацией;
- **Средняя** надежность описывает продукт, который обычно имеет широкий (глобальный) охват и часто обновляется, но предоставляется только одной исходной организацией;
- **Высокая** надежность описывает продукт с глобальным охватом, который часто обновляется и может поставляться несколькими независимыми организациями.

Точность продукта данных (зависит от продукта)

Точность продуктов данных – это оценка неопределенности, связанной с оценками данных. Точность может быть выражена в абсолютном выражении (то есть в физических единицах, таких как мм/год) или в относительном выражении (обычно в процентах). Так, например, если оценка испарения в 200 мм/год имеет ошибку 10% (и, следовательно, точность 90%), реальное значение, вероятно, будет между 180 и 220 мм/год, но может быть ниже, чем 180 или выше 220 мм/год.

Пожалуйста, обратите внимание: редко точность продукта данных НЗ будет столь же высокой, как точность эквивалентного полевого измерения в этом точном месте (хотя это недостаточно часто подвергалось систематическому анализу). При этом в ходе экстраполяции результатов полевых измерений во времени и пространстве НЗ, вероятно, в целом станут более точными.

Область обеспечения качества НЗ находится в стадии быстрого развития, а также разрабатываются стандарты. Мы ссылаемся на деятельность GEO по теме «Стратегия обеспечения качества данных наблюдения Земли: Основные принципы» ([http://wiki.esipfed.org/index.php/Data_Understanding_-_Quality_\(ISO-19157\)](http://wiki.esipfed.org/index.php/Data_Understanding_-_Quality_(ISO-19157))), а также ISO 19157:2013 — Географическая информация – Качество данных для подробной информации.

В публикации Yang и др. (2013 г.) представлена соответствующая сводка по качеству данных при наблюдении Земли. Рис. (ниже) описывает основные аспекты оценки качества данных НЗ.



Рисунок 12.

В своей публикации Yang рассматривает следующие аспекты качества, известные из стандартов ISO, которые можно охарактеризовать как *элементы качества данных производителя*, поскольку они известны производителю данных:

- *полнота*: наличие и отсутствие признаков, их атрибутов и взаимосвязей;
- *логическая согласованность*: степень соблюдения логических правил структуры данных, атрибуция и отношения;
- *позиционная точность*: точность положения объектов;
- *временная точность*: точность временных атрибутов и временных отношений признаков;
- *тематическая точность*: точность количественных признаков и правильность неколичественных атрибутов и классификаций признаков и их взаимосвязей; и
- *преемственность*: информация о происхождении набора данных, включая сведения о применяемой обработке.

Кроме того, Yang и др. добавляют, что записи метаданных хорошего качества также важны. Интересным аспектом является то, что конечные пользователи также хотят получить «приближенные» знания о качестве данных, т. е. комментарии поставщиков данных об общем качестве набора данных, любых ошибках в данных, возможном использовании данных и любой другой информации, которая может помочь оценить пригодность для применения данных. Наконец, большое значение придавалось происхождению набора данных, а также информации о цитировании и лицензировании при оценке того, подходят ли данные для цели.

Несмотря на их обычно более низкую точность, продукты НЗ являются важным источником данных, поскольку изображения НЗ могут предоставлять информацию с большей пространственной протяженностью, пространственной плотностью и/или временной частотой, чем большинство полевых (точечных) сетей наблюдения. Именно по этой причине сочетание НЗ и полевых данных обычно обеспечивает наилучшие информационные результаты. Если возможно иметь модель (например, детерминированную), которая в достаточной мере описывает интересующее явление, то модель НЗ и асимиляции полевых данных является наиболее эффективным способом получения требуемой информации.

Полнота продукта данных (зависит от продукта)

Под зрелостью продукта понимается то, каким образом был создан продукт данных; его можно оценить по уровню проверки, принятия и использования этого продукта данных. Зрелый продукт, как правило, опирается на устоявшуюся науку. Зрелость продуктов данных подтверждается большим количеством рецензируемой литературы, а также соответствующими техническими докладами и прикладными публикациями.

- **Низкая** зрелость указывает на то, что продукт все еще находится на экспериментальной стадии.
- **Средняя** степень зрелости указывает на то, что продукт находится в стадии разработки в том смысле, что, хотя лежащая в его основе идея является зрелой, преобразование продукта в операционный режим все еще продолжается.
- **Высокая** зрелость относится к проверенному, широко протестированному и принятому рабочему продукту.

Сложность продукта данных (зависит от продукта)

Сложность продукта данных описывает уровень сложности, связанный с процессом преобразования данных, обработанных НЗ, в продукт данных. Сложность зависит, среди прочего, от количества задействованных методологических шагов, количества и типа источников входных данных, уровня используемой математики, объемов данных, подлежащих обработке, и/или требуемых технических знаний.

- **Низкая** сложность: отсутствие функции или очень простая функция для преобразования обработанных спутниковых данных в продукт данных, требующий базовых технических знаний;
- **Средняя** сложность: умеренно сложный метод;
- **Высокая** сложность: очень сложный метод создания продукта данных, требующий передовых технических и/или вычислительных знаний.

2.14.2 Необходимость определения «основных потоков данных»: пример GFOI, инициативы GEO

Глобальная инициатива по наблюдению за лесами (GFOI) (<http://www.gfoi.org>) была основана в рамках GEO и в настоящее время возглавляемая правительствами Австралии, Норвегии и США, а также Продовольственной и сельскохозяйственной организацией Объединенных Наций (ФАО) и CEOS. Она пользуется поддержкой широкого сообщества международных экспертов и заинтересованных сторон, в том числе секретариата Рамочной конвенции Организации Объединенных Наций об изменении климата (РКИК ООН), программы кадастров парниковых газов Межправительственной группы экспертов по изменению климата (МГЭИК), Фонда лесного углеродного партнерства Всемирного банка (FCPF), Глобального наблюдения за динамикой лесного и земельного покрова (GOFC-GOLD), университетов, других специализированных международных организаций и экспертов, и, конечно, стран REDD+ (сокращение выбросов в результате обезлесения и деградации лесов в развивающихся странах – СВОД+)

Для таких глобальных инициатив, которые требуют частого, регулярного и глобального охвата, стратегия сбора базовых данных включает координацию основных потоков спутниковых данных, которые удовлетворяют ключевым критериям, согласующимся с принципами реализации инициативы:

- основные потоки данных обеспечивают данные с разрешением 30 м или с более высокой разрешающей способностью, являются бесплатными и доступными для общего пользования для целей GFOI в соответствии с требованиями, предъявляемыми к информации любой страны;
- основные системы потока данных имеют устойчивый и долгосрочный потенциал в плане охвата, обработки и распределения, который соответствует требованиям GFOI в отношении данных в крупном (глобальном) масштабе.

В результате нескольких лет обсуждений в рамках CEOS был достигнут консенсус по рабочему перечню спутниковых миссий CEOS, представляющих основные потоки данных GFOI, на основе имеющейся информации об известной или ожидаемой политике в отношении данных и возможностях миссий. К ним относятся:

Оптические, NIR и TIR системы, требующие ясного неба;

- Landsat-7 и -8 (оптические) – USGS/NASA;
- Серия Sentinel-2 (оптические) – ESA/EU;
- CBERS-3 и -4 (оптические) – INPE (Instituto Nacional de Pesquisas Espaciais, Национальный институт космических исследований Бразилии)/CRESDA (Ресурсный центр спутниковых данных и приложений, Китай);

Миссии радара с синтезированной апертурой, которые имеют способность проникать сквозь облака, туман и дымку;

- Группировка спутников RADARSAT Constellation Mission (радар SAR С-диапазона) – CSA
- Серия Sentinel-1 (радар SAR С-диапазона) – ESA/EU (Европейский Союз);

Миссии, связанные с другими агентствами CEOS, такие как оптические миссии с высоким разрешением (например, RapidEye и SPOT) и радиолокационные миссии (например, ALOS-2,

Radarsat-2, TerraSAR-X и TanDEM -X) могут дополнять основные миссии в районах с постоянной облачностью в дополнение к поддержке проверки и технических исследований.

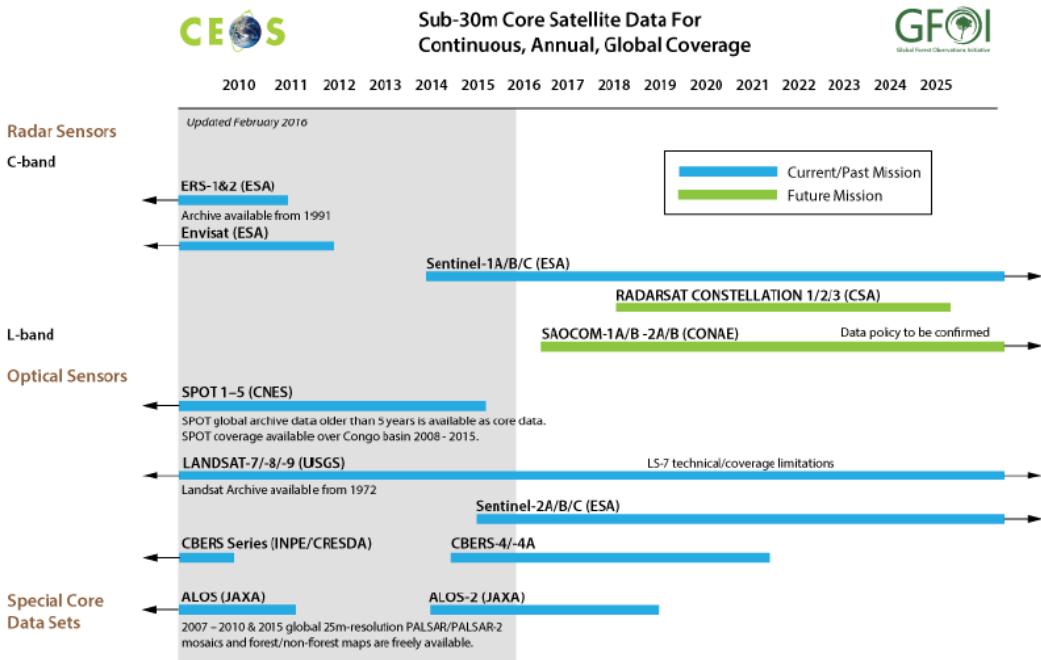
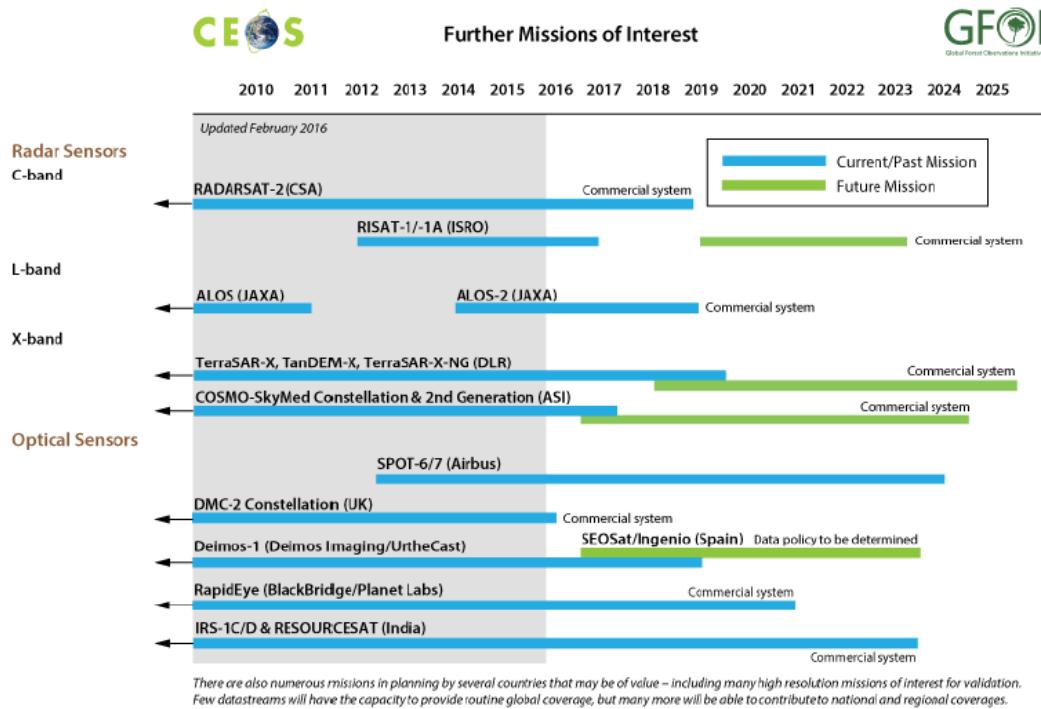


Рисунок 13: Инициатива GFOI: основные спутниковые данные для непрерывного, ежегодного и глобального охвата

2.14.3 Примеры того, как выбирать данные датчиков для измерения конкретных переменных (площадей водного и земельного покрова)

Таблица существующих источников данных НЗ – наиболее «подходящих», по мнению ученых для работы в области НЗ.

В этой таблице показана способность каждого датчика отслеживать и обнаруживать поверхностные воды при наводнениях/паводках. Мы используем разные символы/цвета, чтобы отразить разный потенциал: ①Очень подходит ②Подходит ③Потенциальные ④Не подходит

Таблица 3: АРХИВНЫЕ ИСТОЧНИКИ

Функциональный тип датчика	Инструменты миссии	Краткое название миссии	Пространственное разрешение (m)	Частота (d)	Доступ	Эффективность затрат	Затопления поверхностьюми водами
ОПТИЧЕСКИЙ АКТИВНЫЙ МИКРОВОЛ- НОВЫЙ	TM	Landsat 5	30	16	Открыт	?	②
	MSS	Landsat 1-3	80	18	Открыт	?	②
	A VHRR/2	NOAA 7-14	1100	1	Открыт	?	③
	MODIS	Terra	250	16	Открыт	?	②
	SAR (RADARSAT-1)	RADARSAT-1	25	24	Ограничен	?	②
	PALSAR	ALOS	7-44	14	Ограничен	?	②
	ASAR (stripmap)	Envisat	30		Ограничен	?	②
	ASAR (wide swat)	Envisat	150		Открыт	?	①
	AMI-SAR	ERS-1	30	35	Ограничен	?	②
	AMI-SAR	ERS-2	30	35	Ограничен	?	②
РАДИОЛОКА- ЦИОННАЯ АЛЬТИ- МЕТРИЯ	NRA	TOPEX-Poseidon		10	Открыт	?	④
	GFO-RA	GFO		17	Открыт	?	④
	Poseidon-2	JASON-1		10	Открыт	?	④

Таблица 4: ТЕКУЩИЕ ИСТОЧНИКИ

Функциональный тип датчика	Инструменты миссии	Краткое название миссии	Пространственное разрешение (m)	Частота (d)	Доступ	Эффективность затрат	Затопления поверхностьюми водами
ОПТИЧЕСКИЙ	OSA	IKONOS	3.3		Ограничен	?	②
	GIS	GeoEye	1.6		Ограничен	?	②
	MSI	RapidEye	6.5	1	Ограничен	?	③
	Worldview 2 и 3	Worldview 2 и 3	2/1.3	2	Ограничен	?	③
	MSI	Sentinel-2	10/20/60	5	Открыт	?	③
	ASTER	Terra	15	16	Открыт	?	②
	OLI	Landsat-8	30	16	Открыт	?	②
	ETM+	Landsat-7	30	16	Открыт	?	②
	MODIS	Aqua	250	16	Открыт	?	①
	MODIS	Terra	250	16	Открыт	?	②
	A VHRR/3	NOAA-18	1100	1	Открыт	?	②
	A VHRR/3	NOAA-19	1100	1	Открыт	?	④
	VEGETATION	SPOT-5	1150	26	?	?	④
	VIIIRS	Suomi NPP	1600	16	Открыт	?	④
	OLCI	Sentinel-3 A +B	300	1	Открыт	?	?

Какие датчики следует использовать для измерения параметров растительного покрова (NDVI, Albedo, fPAR и LAI)

Примечание: NDVI (Нормализованный разностный вегетационный индекс)/ fPAR (доля PAR)/ PAR – фотосинтетически активная радиация/индекс площади листьев (LAI).

Формат данных	Функциональный тип датчика	Инструменты миссии	Краткое название миссии	Пространственное разрешение (м)	Период повторного прохождения (дней)	Доступ	Дата запуска	Дата окончания	NDVI	Albedo	fPAR	LAI
Архивный	Оптический	TM	Landsat 5	30	16	Открыт	Июль-82	Июнь-13	1	1	1	1
		MSS	Landsat 1- 3	80	18	Открыт	Июль-72	Сентябрь-83	2	2	2	2
		AVHRR/2	NOAA 7- 14	1100	1	Открыт	Июнь-81		1	2	1	4
Текущий	Активный микроволновый	X-Band SAR	TanDEM- X	16	11	Открыт	Июнь-10		2	4	2	4
		X-Band SAR	TerraSAR-X	16	11	Открыт	Июнь-07		2	4	2	4
		S-Band SAR	HJ-1C	20	31	Открыт	Ноябрь-12		2	4	2	4
		SAR (RADARSAT -2)	RADARSAT -2	25	24	Ограничено	Декабрь-07		2	4	2	4
	Оптический	MSI	RapidEye	6.5	1	Ограничено	Август-08		2	2	2	2
		MSI	Sentinel-2	10/20/60	5	Открыт	Июнь-15		2	2	2	2
		ASTER	Terra	15	16	Открыт	Декабрь-99		1	1	1	1
		OLI	Landsat-8	30	16	Открыт	Февраль-13		1	1	1	2
		ETM+	Landsat 7	30	16	Открыт	Апрель-99		1	1	1	2
		Hyperion	NMP EO-1	30	16	Открыт	Ноябрь-00	Октябрь-14	1	1	1	1
Архивный	AWiFS	RESOUR CESAT-2	55	26	Открыт	Апрель-11			1	1	1	1
	LISS-III (Resourcesat)	RESOUR CESAT-2	55	26	Открыт	Апрель-11			1	1	2	2
	MISR	Terra	250	16	Открыт	Декабрь-99			2	2	2	2
	MODIS	Aqua	250	16	Открыт	Май-02			1	1	1	1
	MODIS	Terra	250	16	Открыт	Декабрь-99			1	1	1	1
	AVHRR/3	NOAA-18	1100	1	Открыт	Май-05			1	2	1	4
	AVHRR/3	NOAA-19	1100	1	Открыт	Февраль-09			1	2	1	4
	VEGETATION	SPOT-5	1150	26	?	Май-02	Декабрь-14		1	2	1	2

		VIIRS	Suomi NPP	1600	16	Открыт	Октябрь-11	1	1	1	1
Будущий	Активный микроволновый	SAR (RCM)	RADARS AT C-1	50	12	Ограничен	2018	2	4	2	4
		SAR (RCM)	RADARS AT C-2	50	12	Ограничен	2018	2	4	2	4
		SAR (RCM)	RADARS AT C-3	50	12	Ограничен	2018	2	4	2	4
	Оптический	LISS-III (Resourcesat)	RESOUR CESAT-2A	5.8	26	Открыт	2015	1	1	2	2
		MSI (Sentinel-2)	Sentinel-2A	10	10	Открыт	2015	1	2	1	2
	MSI (Sentinel-2)	Sentinel-2B	Sentinel-2B	10	10	Открыт	2016	1	2	1	2
		MSI (Sentinel-2)	Sentinel-2C	10	10	Открыт	2020	1	2	1	2
	LISS-IV	RESOUR CESAT-2A	23.5	26	Открыт	2015	1	1	2	2	
		HISUI	ALOS-3	30	60	?	2016	1	1	1	1
	AWiFS	RESOUR CESAT-2A	55	26	Открыт	2015	1	1	1	1	
		OLCI	Sentinel-2A	300	27	Открыт	2015/16	1	2	2	2
		OLCI	Sentinel-2B	300	27	Открыт	2017	1	2	2	2
	OLCI	Sentinel-2C	Sentinel-2C	300	27	Открыт	2020	1	2	2	2
		VIIRS	JPSS-1	1600	16	Открыт	2017	1	1	1	1

Источник: 1. Committee on Earth Observation Satellites (CEOS) Earth Observation Handbook (<http://www.eohandbook.com/>) 2. WMO Observing Systems Capability Analysis and Review Tool (<http://www.wmosat.info/oscar/>).

2.15 Государственные и частные датчики/спутники

Есть три типа поставщиков спутниковых данных; национальные и многонациональные космические агентства, большинство из которых являются частью CEOS, коммерческие частные отраслевые поставщики и некоторые частные поставщики, связанные с национальными космическими агентствами, такие как ISRO для Индии и ее дочерняя компания Antrix.

В таблицах 3 и 4 упоминаются «открытые» и «ограниченные» источники спутниковых данных. Большинство открытых источников поступает от национальных и многонациональных космических агентств, а наиболее ограниченными являются коммерческие поставщики. Причина, по которой они называются ограниченными, заключается в том, что Лицензионное Соглашение Конечного Пользователя обычно довольно ограничивает обмен исходными данными. Многие коммерческие поставщики спутниковых данных уделяют особое внимание весьма высокому пространственному разрешению для целей обороны, точного земледелия и точного картирования (т.е. кадастрового).

Большинство источников НЗ, используемых для мониторинга переменных экосистем или возобновляемых и невозобновляемых ресурсов, поступают от космических агентств с политикой открытых данных, поскольку это позволяет сократить или аннулировать расходы на получение данных. Однако для мониторинга дорогостоящих видов сырья действительно используются имеющиеся в продаже данные с очень высоким пространственным разрешением.

Хотя в прошлом большая часть НИОКР проводилась космическими агентствами или крупными технологическими институтами и академическими кругами, в последние несколько лет наблюдается расширение прикладных исследований такими организациями, как GEO, НПО, международные банки развития и частный сектор. Таким образом, крайне важно придерживаться прагматичного подхода при вступлении в эпоху больших данных и работать со всеми секторами. Все больше компаний начинают запускать свои собственные спутники или использовать продукты спутниковых данных, а некоторые берут на себя инициативу в отдельных областях.

Даже если раньше правительства полагались только на свои национальные спутники и датчики, теперь они понимают острую необходимость вести переговоры и работать с частными компаниями, поскольку последние могут предлагать услуги спутниковых данных или продукты, которые могут быть полезны для управления большими данными и предлагать решения. В этом разделе приведены некоторые примеры частных источников данных НЗ, однако этот перечень ни в коей мере не является исчерпывающим.

Растет число компаний, которые инвестируют как в спутники, так и в получаемые данные, хотя также наблюдается волна консолидации, сопровождающаяся слиянием компаний. В то же время быстро развиваются такие технологические ускорители, как наноспутники, кубсаты, микроспутники и т. д., и расширяют возможности космических агентств и компаний, которые запускают и эксплуатируют крупные системы с длительным сроком службы. Таким образом, в настоящее время существует масштаб деятельности по созданию и запуску спутников весом от 100 до 1000 килограммов с длительным сроком службы и очень точными измерениями для малых или кубических спутников, которые весят от одного до нескольких килограммов, но имеют ограниченный срок службы.

В этой главе содержится информация об имеющихся веб-ресурсах для множества доступных спутников и некоторые руководящие принципы выбора подходящего спутника для данной работы. Базы данных CEOS и OSCAR больше ориентированы на системы государственного сектора.

Некоторые частные компании, которые предоставляют весьма актуальные услуги, представлены здесь, но только в виде краткого описания и без намерения быть исчерпывающим. Некоторые более крупные компании, имеющие опыт предоставления наборов данных с высоким пространственным разрешением и/или глобальных данных:

- Digital Globe (<https://www.digitalglobe.com/>)
- AIRBUS Industries (<http://www.satimagingcorp.com/satellite-sensors/other-satellite-sensors>)
- Planet (<https://www.planet.com/>)
- Google Terra Bella (<https://terrabella.google.com>)

Существует множество небольших специализированных компаний, которые могут иметь свои собственные спутники, такие как гибрид космического агентства и промышленности; Китайско-бразильская спутниковая программа изучения ресурсов Земли (CBERS) и сотрудничество между ISRO и ANTRIX в Индии. Тенденция заключается в том, что космические агентства все чаще создают совместные предприятия с частными компаниями.

Политика открытых данных приветствуется, когда речь заходит о государственных структурах. Частные поставщики спутниковых данных, как правило, не придерживаются такой политики, потому что они могут получить прибыль от продажи таких данных. Однако, в этом секторе все может измениться. В сентябре 2015 года Уилл Маршалл, генеральный директор Planet, объявил, что компания откроет доступ (в соответствии с открытой лицензией CC BY SA 4.0) к некоторым спутниковым данным высокого разрешения, полученным ее группировкой микроспутников. Маршалл сделал заявление на Ассамблее ООН, рассказав о том, как данные будут способствовать достижению ЦУР ООН. Другими словами, их данные можно использовать бесплатно, но все, что производится с их использованием, должно быть сделано аналогичным образом открытым (и это относится и к продуктам с добавленной стоимостью)¹³.

Компания Digital Globe – ведущий мировой поставщик коммерческих продуктов и услуг в области изображений Земли с высоким разрешением в США. Она активировала сервис First Look, службу подписки, которая обеспечивает управление чрезвычайными ситуациями и гуманитарных работников быстрым веб-доступом к изображениям пострадавшего района до и после события. Digital Globe предоставляет изображения для помощи в гуманитарном кризисе через сервис Tomnod, еще один проект, использующий краудсорсинг для идентификации объектов и мест на спутниковых снимках.

Поэтому эти компании признают, что, хотя спутниковые изображения полезны сами по себе, большую пользу приносит извлечение значимой информации, которая может быть использована службами экстренного реагирования и спасательными службами.

Другой тип подхода используется компаниями, которые вообще не имеют космических активов, но собирают, используют и отображают данные о Земле от реальных поставщиков, таких как Google Earth Engine, Amazon Web Services, Facebook и т. д. Проект Facebook Internet.org (карта населения) представляет собой пример, целью которого является передача доступа в Интернет в районы мира, где отсутствует широкополосный доступ, с помощью дронов и спутников. Этот доступ к Интернету может затем использоваться для обработки данных НЗ, обеспечивающих геопространственную информацию.

2.16 Датчики будущего и архитектуры данных

Базы данных датчиков Всемирной метеорологической организации OSCAR и CEOS также включают всесторонний перечень будущих датчиков НЗ.

В целом датчики НЗ становятся все более совершенными в связи с технологическим развитием фактических датчиков на спутниковых платформах, развитием самих платформ, электроники, сжатия на борту и скорости передачи на наземные станции и т. д. Эти разработки также ведут к миниатюризации. Миниатюризация приводит либо к расширению возможностей на спутниках

¹³ http://www.eurisy.org/article-free-and-open-satellite-data-private-companies-join-in-the-game_18#sthash.0VQ20Ubf.dpuf

одинакового размера (обычно 500-1000 кг), либо на спутниках меньшего размера, вплоть до наноспутников, или на кубсатах весом всего несколько килограммов.

Еще одним новшеством в области датчиков LEO является гонка между публично финансируемыми и общедоступными потоками данных НЗ и коммерческими спутниковыми датчиками НЗ. Как правило, датчики с низким разрешением (размер пикселя более 20 м) предоставляются бесплатно большинством космических агентств, а данные с более высоким разрешением предоставляются по коммерческим ценам. Однако происходит постепенный сдвиг в общедоступном пространственном, временном и спектральном разрешении: например, до появления Sentinel-2 в 2015 году со свободно доступным пространственным разрешением 10 м, Landsat с размером пикселей равным 30 м был бесплатным источником данных с самым высоким разрешением. Таким образом, коммерческие операторы теперь должны предоставлять данные разрешением менее 10 м на пиксель, чтобы быть конкурентоспособными. В настоящее время самым высоким пространственным разрешением, предлагаемым коммерческой компанией, являются панхроматические данные WorldView-3 0,4 м и мультиспектральные и SWIR-данные 1,3 м.

Другими областями конкуренции являются временное разрешение и охват территорий. Представляя изображения на основе принципа «от края до края», Landsat требуется 16 дней, в то время как Sentinel-2 (после запуска обоих спутников 2A и 2B) будет предоставлять изображения примерно один раз в 5 дней по всему миру. Спектральное разрешение также увеличивается, что приведет к переходу от мультиспектральных датчиков (обычно с 4–13 спектральными каналами) к спектрометрам формирования изображения (с сотнями спектральных каналов). Коммерческие операторы обеспечивают более высокое временное разрешение, либо путем наведения своих датчиков НЗ во всех направлениях в пространстве, либо путем запуска «роев» датчиков, таких как спутниковая группировка RapidEye, а также все более широкое распространение получают кубсаты и наноспутники, такие, как спутниковая группировка SkySat, состоящая из 7 спутников высокого разрешения в дополнение к существующему флоту Planet из 60 спутников среднего разрешения. Первая группировка обеспечивает регулярные, быстро обновляемые снимки отдельных районов земного шара с разрешением менее метра; последняя – регулярный глобальный охват с разрешением 3–5 метров.

Геостационарные датчики, которые были исключительно областью крупномасштабных метеорологических измерений, теперь обеспечивают изображения с более высоким пространственным разрешением и большим количеством спектральных каналов, особенно с добавлением синего и зеленого каналов, что делает их гораздо более полезными для неметеорологического использования. Термальные каналы также становятся более точными в разрешении. Самым последним типом усовершенствованного датчика GEO является спутник Himawari-8 из Японии, который расположен над Индонезией на высоте 36 000 км. Он просматривает половину земного шара каждые 10 минут с пространственным разрешением от 500 до 2000 м на пиксель. Еще несколько датчиков типа Himawari (известных как серия GOES-R, первый запуск которых состоялся в конце 2016 года для Северной и Южной Америки) будут размещены на геостационарных орбитах для охвата Америки, Африки и Европы. Наблюдается быстрый прогресс в развитии неметеорологического использования этого передового типа геостационарных данных НЗ.

Потенциальные новые архитектуры данных также имеют решающее значение для будущего НЗ и преимуществ использования данных. Программы космических агентств по исследованию Земли сталкиваются с рядом тенденций, которые вызывают необходимость изменения способов обработки, анализа и распространения данных. Поскольку это изменение приобретает решающее значение, председатель CEOS (CSIRO) на 2016 год предложил изучить проблемы, связанные с архитектурами систем данных НЗ следующего поколения (специальная группа CEOS Future Data Architecture (FDA)) и возможности, предлагаемые подходами, такими как Data Cube (см. 2.17), а также коммерческие облачные решения для хранения и обработки данных. Специальная группа рассмотрела соответствующие текущие инициативы, предпринятые CEOS; уроки, извлеченные из первых прототипов, определила основные проблемы и сформулировала некоторые рекомендации.

Промежуточный отчет теперь доступен на веб-сайте CEOS¹⁴. Более полный отчет будет представлен в 2017 году.

Максимизация ценности НЗ фактически является одним из основных факторов, определяющих деятельность космических агентств. В последние годы во всех космических агентствах наблюдается устойчивая тенденция к более широкой интеграции различных данных НЗ для различных видов применения в решении земельных, водных и климатических вопросов. Усиление интеграции привело к появлению более широкого спектра применений, что усложняет цепочку создания стоимости, когда наблюдения сочетаются с аналитикой для удовлетворения многочисленных потребностей пользователей.

Еще одна важная идея для НСУ заключается в том, что космические агентства, помимо бесплатных и открытых данных, осознают, насколько желательно сейчас получить бесплатный и открытый доступ к аналитическому программному обеспечению и инструментам, которые облегчают использование данных НЗ. Инициатива CEOS Data Cube, основанная на «Австралийском кубе геонаучных данных» (AGDC), см. раздел 2.17, является примером этой тенденции. Проект опирается на программное обеспечение с открытым исходным кодом для создания кубов данных (прием спутниковых данных) и взаимодействия с кубами данных (интерфейсы прикладного программирования = API). Программное обеспечение с открытым исходным кодом, несомненно, будет стимулировать инновации в области приложений и более широкое использование спутниковых данных, поскольку эти передовые технологии могут использоваться во всем мире и даже в развивающихся странах, которые не являются традиционными пользователями спутниковых данных.

В отчете FDA подчеркивается, что агентствам CEOS необходимо будет продолжить работу над этим вопросом, учитывая их интересы. «Будущая работа должна помочь пользователям извлечь выгоду из всех соответствующих данных CEOS на всех этапах – дополняя прошлые усилия, которые подчеркивали единство «обнаружения», без полного решения значительных проблем в анализе и использовании разрозненных или несовместимых данных после обнаружения».

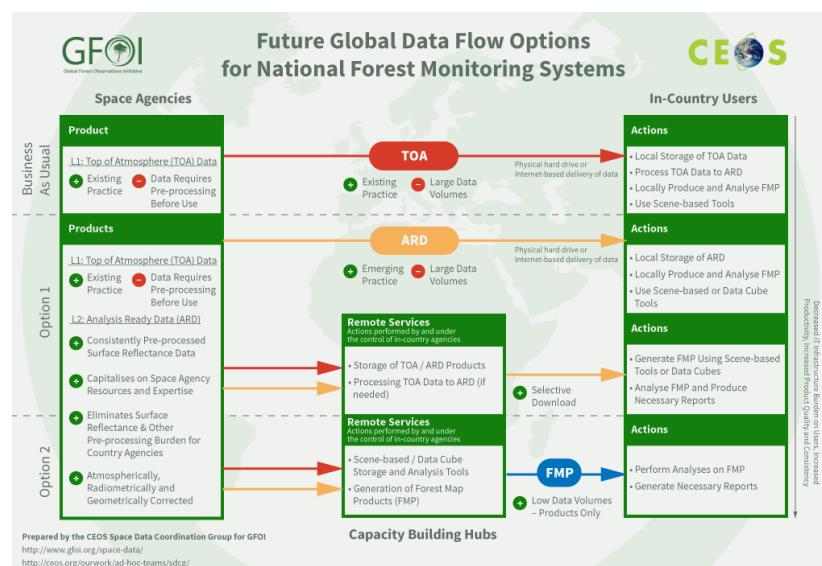


Рисунок 14: Будущие варианты глобального потока данных для национальных систем мониторинга лесов

¹⁴ Исследование будущих архитектур доступа к данным и анализа CEOS – промежуточный отчет (октябрь 2016 г.):

<http://ceos.org/meetings/30th-ceos-plenary/> и более подробно по ссылке:

http://ceos.org/document_management/Meetings/Plenary/30/Documents/5.2_Future-Data-Architectures-Interim-Report_v.1.pdf

2.17 Пример: инициативы Data Cube Австралийский куб геонаучных данных

Австралийский куб геонаучных данных (AGDC) – это новый совместный подход к хранению, организации и анализу огромного количества спутниковых изображений и других данных наблюдений Земли, позволяющий быстрее и проще предоставлять информацию по вопросам, которые могут затронуть всех австралийцев. AGDC разработан и поддерживается партнерством между Geoscience Australia, Государственным объединением научных и прикладных исследований (CSIRO) и Национальной вычислительной инфраструктурой Австралии (NCI).

AGDC делает стандартизированные и откалиброванные спутниковые данные доступными в структурах на различных платформах обработки, которые могут быть локальными, удаленными или облачными. Программное обеспечение AGDC также оптимизировано для использования средствами высокопроизводительных вычислений, такими как средство Raijin в Национальной вычислительной инфраструктуре в Австралии. AGDC состоит из ряда структур данных и инструментов, которые облегчают эффективный анализ очень больших коллекций стандартизированных и калиброванных данных с привязкой к координатной сетке, таких как спутниковые данные для различных целей (рис. 15). Это позволяет пользователю легко применять весь потенциал спутниковых данных для решения реальных задач, таких как производительность сельского хозяйства и мониторинг промышленного воздействия на ландшафт.

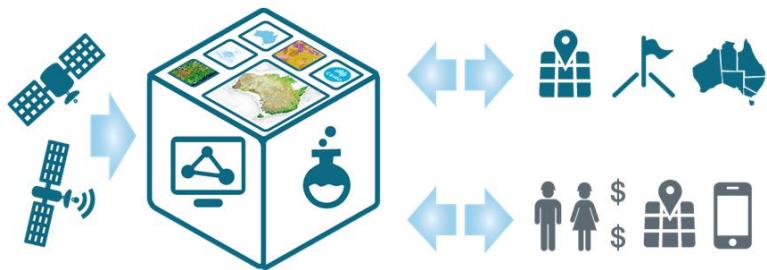


Рисунок 15: Куб данных обеспечивает общую аналитическую основу, позволяющую нескольким источникам данных (слева) создавать информацию для многократного использования (справа).

Стандартизация и калибровка повышают значение, которое может быть получено на основе НЗ и других источников больших наборов данных с координатной привязкой. Они позволяют быстро разрабатывать информационные продукты, с тем чтобы правительства и промышленность могли принимать обоснованные решения.

Такая стандартизация и калибровка важны и по другой причине. Это означает, что продукты, созданные разными пользователями для решения различных проблем, могут быть интегрированы. Такая способность к интеграции информации будет иметь решающее значение для решения крупных национальных задач. Например, усилия по повышению производительности сельского хозяйства являются более эффективными, если информация о влажности почвы может быть интегрирована с информацией о здоровье сельскохозяйственных культур и о прогнозируемых осадках.

Доказательство концепции

AGDC использовался для выполнения анализа, который ранее был непрактичным. Например, недавно завершенное исследование позволило картировать наблюдения поверхностных вод на всей территории Австралии за период 1998-2012 годов с номинальным разрешением 25 метров. При использовании традиционной модели этот анализ занял бы 4-8 лет, но теперь его можно провести менее чем за 4 часа с использованием AGDC. Этот продукт теперь является источником информации для Национального информационного портала о рисках наводнений (рис. 16), помогая населению понимать и планировать возможные будущие наводнения.

Участок суши на Рисунке 16 показывает область канала, прилегающую к пустыне Симпсон (песчаная пустыня в центре Австралии), с редко наблюдаемой водой (красным цветом) и постоянными водоемами (синим цветом).



Рисунок 16: Наблюдения за водой из космоса (WOFS) позволили нанести на карту наблюдения за поверхностными водами для всей Австралии с разрешением 25 метров, используя данные с 1998 по 2012 год.

Ключевые характеристики

Использование стандартизованных и калиброванных данных является ключевой характеристикой AGDC. В прошлом калибровка данных по НЗ производилась на индивидуальной основе для подмножеств данных. Это может быть неэффективным и приводить к большим накладным расходам. Путем калибровки всех данных в соответствии с одним и тем же стандартом, он сразу же становится пригодным для использования многими различными группами пользователей, устранив до 80% непродуктивной «начальной деятельности», часто необходимой в традиционном подходе.

Еще одной ключевой характеристикой куба данных является то, что он поддерживает каждое уникальное наблюдение и делает его доступным для анализа. Это контрастирует с традиционными подходами, где технические ограничения означают, что используются только небольшие подмножества. Это также позволяет использовать в работе полный «временной ряд», все наблюдения, полученные спутниками за многие годы. Такой подход очень важен, потому что многие важные явления становятся очевидными только при регулярном наблюдении с течением времени.

Последней ключевой характеристикой AGDC является его способность обеспечивать «экономию за счет масштаба». Одни и те же дорогостоящие суперкомпьютеры и хранилища данных, инфраструктура, которая позволяет достичь таких уровней производительности, могут использоваться для различных целей. Способность реализовать эту экономию за счет масштаба приобретает все большее значение.

Объемы данных, которые уже и без того невозможно обрабатывать на индивидуальной основе, будут расти в геометрической прогрессии, поскольку новые спутники из США, Европы, Японии и других стран предоставят нам доступ к все большему количеству данных, чем когда-либо ранее. Подход, позволяющий инвестировать в общую инфраструктуру, будет иметь решающее значение для получения выгод от этих данных.

Уникальный способ, с помощью которого AGDC организует данные, позволяет быстро манипулировать ими в одной высокопроизводительной вычислительной среде (Рисунок 17) для

многих различных целей. AGDC организует изображения из нескольких источников и разного времени, например, изображения слева, в стопки (справа). Вершиной стека являются самые последние наблюдения. В результате AGDC может помочь нам справиться с «потоком данных».

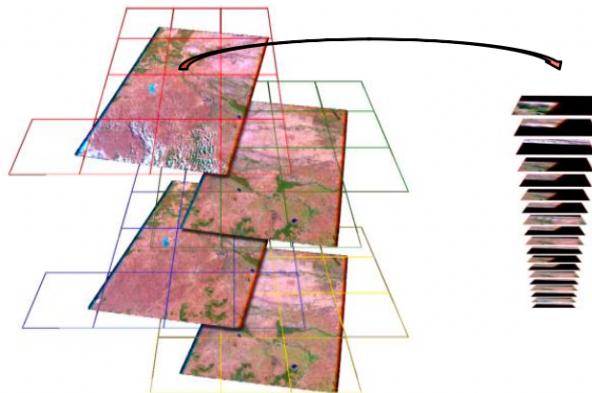


Рисунок 17: Процесс организации AGDC

Дополнительную информацию можно найти на сайте www.datacube.org.au, а исходный код доступен на GitHub по адресу <https://github.com/data-cube/agdc-v2>.

Роль CEOS и технологии Data Cube

Многим странам не хватает знаний, инфраструктуры и ресурсов для доступа и использования космических данных. Страны выразили пожелание, чтобы CEOS оказывал им поддержку путем предоставления инструментов поиска и обнаружения, облачного хранения и обработки данных, а также подготовки кадров и наращивания потенциала.

CEOS возглавляет несколько проектов по демонстрации инструментов обслуживания данных и созданию архитектуры, поддерживающей растущий объем данных. Некоторые страны (Кения и Колумбия) уже обратились к CEOS с просьбой помочь им в создании такой среды для управления большими данными НЗ и активизации пилотных проектов для Data Cube. Цель состоит в том, чтобы эти пилотные проекты заложили прочную основу для будущих операционных систем, которые будут финансироваться и управляться организациями ООН и отдельными странами.

Обеспечивая для стран архитектуру и средства доступа к спутниковым данным и их использования, а также подготовку кадров и наращивание потенциала, необходимых для использования этих средств и управления спутниковыми данными, CEOS разрабатывает прототипы проектов для демонстрации этих инструментов и услуг. CEOS также оценивает передовые технологии в своей архитектуре и инструментах, таких как платформы служб данных и кубы данных на основе облачных вычислений.

Страны и регионы с текущими и будущими проектами Data Cube:

- Австралия
- Кения
- Колумбия
- Швейцария
- Юго-Восточная Азия

Национальным статистическим службам, заинтересованным в этих будущих программах данных, настоятельно рекомендуется связаться с отделом системного проектирования CeOS, поскольку они создают глобальную систему для организации внедрения кубов данных в странах.

2.18 Рекомендации по выбору соответствующих источников НЗ

По мере того как НЗ переходит от использования нескольких общедоступных и четко определенных датчиков НЗ, таких как источники MODIS и Landsat, к гораздо более широкому набору датчиков со значительно улучшенными характеристиками, тенденция будет заключаться в увеличении объема данных, с большей точностью, будучи частью новой эры больших данных.

Прежде чем использовать спутниковые данные НЗ, НСУ должны рассмотреть некоторые ключевые вопросы (см. схему решений на рисунке 33, глава 5), чтобы решить, будет ли использование НЗ способствовать эффективному достижению их целей и результатов. Исследовательский вопрос может помочь пользователю решить, какие источники данных НЗ могут быть подходящими и полезными.

Если данные НЗ выбраны в качестве источника данных, рассмотрите возможность использования оптимальных датчиков (см. базы данных CeOS или OSCAR BMO) в зависимости от различных ранее определенных критериев. Глобальная инициатива GFOI является хорошим примером мониторинга земной поверхности в качестве основного потока данных, который был четко определен и в настоящее время действует.

Даже несмотря на то, что точность для одного пикселя, как правило, ниже, чем при полевых измерениях для одного и того же участка, продукты НЗ по-прежнему следует рассматривать в качестве важного альтернативного источника данных, поскольку снимки НЗ могут предоставлять информацию в более широких пространственных масштабах, пространственной плотности и/или временной частоте по сравнению с любыми полевыми (точечными) сетями наблюдения. Расходы на единицу обследуемой площади и на переменную значительно сократятся по мере распространения спутниковых данных НЗ.

Благодаря использованию интеллектуальных методов для извлечения все более достоверной информации из данных НЗ, акцент может все больше смещаться с полевых данных на данные НЗ. НСУ могут извлечь пользу из использования большего объема данных НЗ, когда это целесообразно и уместно, для замены или дополнения своих результатов и анализа.

Еще один момент, на который следует обратить внимание – это важность общедоступных данных, предоставляемых всеми национальными космическими агентствами. Все больше и больше частных компаний инвестируют в спутники и запускают свои собственные миссии, чтобы реагировать на немедленные, своевременные и оперативные запросы своих пользователей. Несмотря на большой интерес к этим новым видам деятельности в космосе, здесь важно подчеркнуть глобальный охват и преемственность во времени (миссии, запланированные на следующие десятилетия); две основные характеристики, которые до настоящего времени предоставлялись только государственными и национальными космическими агентствами.

Этот фундаментальный аспект, безусловно, необходимо учитывать НСУ, если они хотят использовать спутниковые данные НЗ в определенных ситуациях. Например, если область интереса является то, как соответствующий процесс, такой как эрозия, деградация земель, типы сельскохозяйственных культур, лесной покров, протяженность водно-болотных угодий, эвапотранспирация (суммарное испарение), расширение городов – развивался за последние 30 лет, очевидно, что временной ряд Landsat единственный источник данных. Однако, например, если интересующий вопрос заключается в том, что недавняя добыча газа из угольных пластов вызвала какое-либо проседание, то наиболее подходящими будут общедоступные или частные SAR-изображения высокого разрешения в интерферометрическом режиме. Эти источники дадут точность миллиметровой деформации в любом временном масштабе, полученную с момента запуска спутников. Если требуется базовая геопространственная карта окружающей среды или

природных ресурсов, лучше всего выбрать изображения с самым высоким доступным пространственным разрешением.

С концепцией ARD глобальный охват и непрерывность являются важными вопросами, которые требуют значительного количества ресурсов для создания куба данных НЗ типа ARD. Это имеет смысл только в том случае, если данные о прошлом, настоящем и будущем будут многократно использоваться.

2.19 Иллюстрации

ООН определила 17 целей в области устойчивого развития (ЦУР) для решения проблем экономического и социального развития и экологической устойчивости. Для каждой из ЦУР ООН было определено несколько задач и разрабатываются показатели, которые страны будут использовать для отслеживания прогресса и отчетности. Для достижения ЦУР геопространственная информация и данные наблюдений Земли, собранные на местном, национальном и глобальном уровнях и подкрепленные передовыми научными данными, инструментами и технологиями, могут сыграть важную и полезную роль в мониторинге целей, отслеживании прогресса и оказании помощи странам в корректировке промежуточных результатов. В сочетании с демографическими и статистическими данными эти источники помогают странам анализировать и моделировать условия, создавать карты и отслеживать изменения с течением времени последовательным и стандартизованным образом (<http://www.earthobservations.org/>).

Ниже проиллюстрированы 2 ЦУР, которые имеют идентифицированный измеримый компонент наблюдения Земли.

Примеры использования спутниковых данных НЗ для целей ООН в области устойчивого развития (Повестка дня на период до 2030 г.)

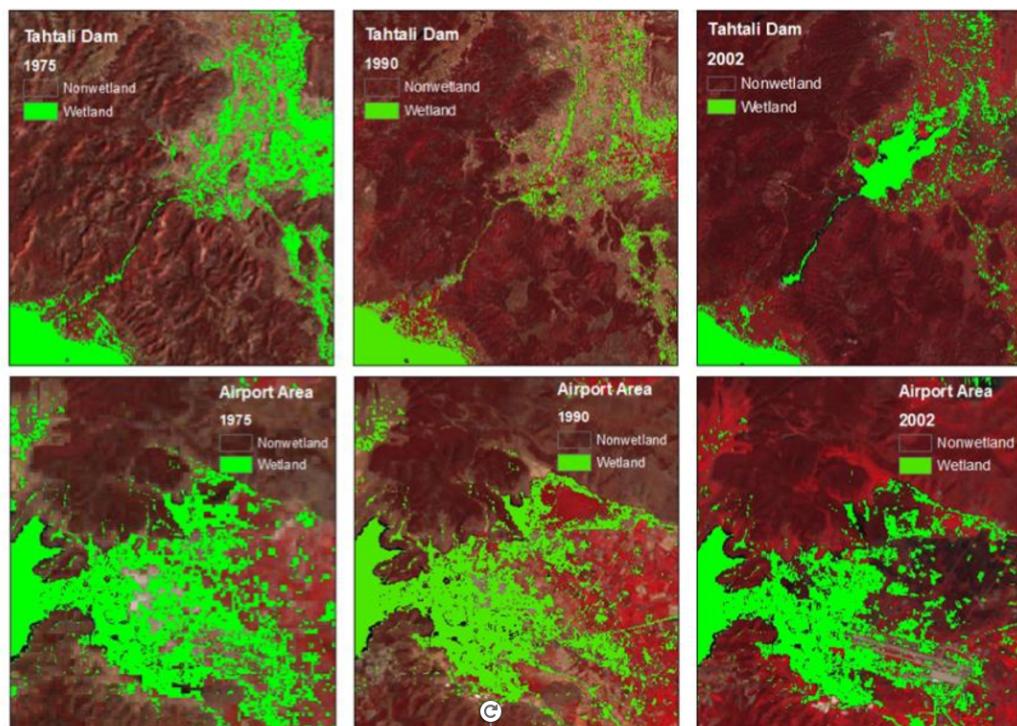


6 ЧИСТАЯ ВОДА И САНИТАРИЯ



Разновременные продукты идентификации (Landsat 1975 г., 1990 г. и 2002 г.) и разграничения водно-болотных угодий для типовых участков между Измиром и Бодрумом (верхняя часть: район вокруг плотины Тахталы; нижняя часть: район аэропорта Бодрума).

<http://www.earthzine.org/wp-content/uploads/2011/12/Figure-3.jpg>



13 CLIMATE ACTION



14 LIFE BELOW WATER



15 СОХРАНЕНИЕ ЭКОСИСТЕМ СУШИ



16 PEACE, JUSTICE AND STRONG INSTITUTIONS



17 PARTNERSHIPS FOR THE GOALS

Цель 15
«Защита и восстановление экосистем суши и содействие их рациональному использованию, рациональное лесопользование, борьба с опустыниванием, прекращение и обращение вспять процесса деградации земель и прекращение процесса утраты биоразнообразия»

Цель 15.2
«К 2020 году содействовать внедрению методов рационального использования всех типов лесов, остановить обезлесение, восстановить деградировавшие леса и значительно расширить масштабы лесонасаждения и лесовосстановления во всем мире»

Показатель 15.2.2
Чистая безвозвратная потеря леса

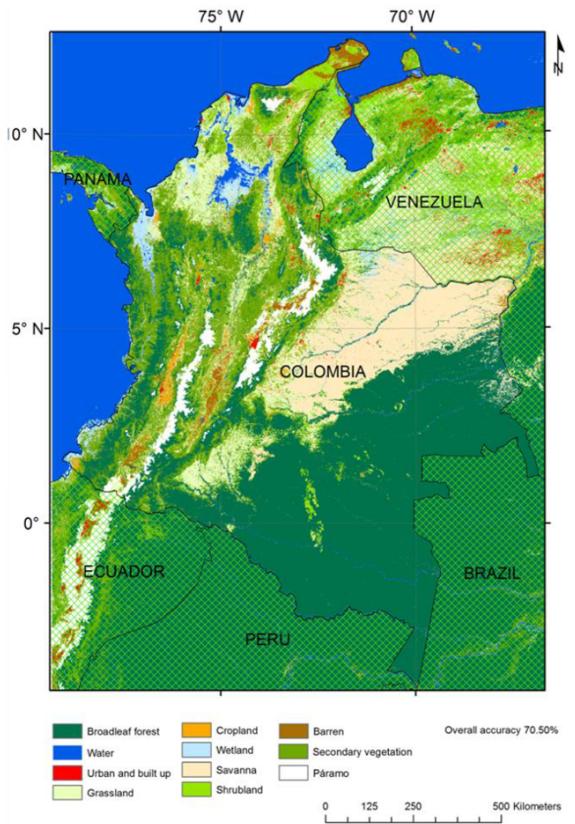
Земной покров (временные ряды), подлежащий мониторингу, например, глобальная инициатива GFOI (для обеспечения устойчивой доступности наблюдений для национальных систем мониторинга лесов)





Карта земного покрова Колумбии с использованием данных MODIS NDVI за 2011 г. +/- 2 года и данных об отражательной способности поверхности MOD12A1, отфильтрованных для некачественных наблюдений и добавленных высот (TS-F-C2-E)

<http://www.mdpi.com/2072-4292/7/12/15833/htm>





	Распределение населения	Картирование городов и инфраструктуры	Высоты и рельеф	Картирование земного покрова и землепользования	Океанографические наблюдения	Наблюдения в области гидрологии и качества воды	Мониторинг атмосферы и качества воздуха	Наблюдения биоразнообразия и экосистем	Сельскохозяйственный мониторинг	Мониторинг опасностей, катастроф и влияния окружающей среды
1 Ликвидация нищеты										
2 Ликвидация голода										
3 Хорошее здоровье и благополучие										
4 Качественное образование										
5 Гендерное равенство										
6 Чистая вода и санитария										
7 Недорогостоящая и чистая энергия										
8 Достойная работа и экономический рост										
9 Индустриализация, инновации и инфраструктура										
10 Уменьшение неравенства										
11 Устойчивые города и населенные пункты										
12 Ответственное потребление и производство										
13 Борьба с изменением климата										
14 Сохранение морских экосистем										
15 Сохранение экосистем суши										
16 Мир, правосудие и эффективные институты										
17 Партнерство в интересах устойчивого развития										

Данные ЦУР и НЗ: спутниковые данные, используемые в различных областях применения (документ GEO)¹⁵

¹⁵ GEO «Поддержка официальной статистики в мониторинге ЦУР», март 2016 г. – 47-я сессия Статистической комиссии ООН в Нью-Йорке. Подробнее здесь: http://www.earthobservations.org/documents/meetings/201603_eo_sdgs_ny/2016_geo_un_flyer.pdf, а также здесь: http://www.earthobservations.org/geo_sdgs.php

**Цели в области устойчивого развития
Наблюдения Земли на службе Повестки дня на период до 2030 года**

Цель										Цель	Показатель				
Способствовать прогрессу в достижении Цели, но не показателя как такового											Прямая мера или косвенная поддержка				
							1,4	1,5			1.4.2				
					2.3	2,4	2.c				2.4.1				
					3.3	3.4	3,9	3.д			3.9.1				
								5.a			5.a.1				
		6.1	6.3	6.4	6,5	6,6	6.a	6.6			6.3.1	6.3.2	6.4.2	6.5.1	6.6.1
					7.2	7.3	7.a	7.6			7.1.1				
								8.4							
					9.1	9.4	9,5	9.a			9.1.1	9.4.1			
						10,6	10,7	10.a							
	11.1	11.3	11,4	11,5	11,6	11,7	11.6	11.c			11.1.1	11.2.1	11.3.1	11.6.2	11.7.1
				12.2	12,4	12,8	12.a	12.6			12.a.1				
					13.1	13.2	13.3	13.6			13.1.1				
		14.1	14.2	14.3	14,4	14,6	14,7	14.a			14.3.1	14.4.1	14.5.1		
	15.1	15.2	15,3	15,4	15,5	15,7	15,8	15,9			15.1.1	15.2.1	15.3.1	15.4.1	15.4.2
								16,8							
17.2	17,3	17,6	17,7	17,8	17,9	17.16	17.17	17.18			17.6.1	17.18.1			

Источник: Инициатива CEOS/GEO EO4EDG, 2017 г.

3. Методология

3.1 Введение

В этой главе представлен обзор методологии использования спектральных каналов или частот, полученных из данных наблюдения Земли, для оценки величин, представляющих интерес для НСУ. Эта глава следует за главой 2, в которой описываются источники данных НЗ, и рассматривается вопрос: «Теперь, когда у нас есть данные, как мы можем их использовать?».

Изображения НЗ и данные, извлекаемые из них, попадают в категорию «больших данных». Этот термин обычно используется для описания больших наборов данных, которые сложно хранить, обрабатывать и анализировать. Для многих НСУ традиционные инструменты управления и анализа данных неадекватны. Другие вопросы включают выбор соответствующих статистических данных, рассмотрение качества и точности данных и результатов, а также проверку и интерпретацию выходных данных.

Разработка методов и инструментов для анализа данных НЗ ведется уже более десяти лет. Однако увеличение разрешающей способности и доступности данных НЗ означает, что в настоящее время такие организации и учреждения, как НСУ, могут рассмотреть вопрос о включении этого источника данных в свои системы.

Эта глава начинается в разделе 3.2 с обсуждения типов оценок, относящихся к НСУ, которые могут быть получены на основе данных НЗ. К ним относятся официальная статистика для использования на национальном уровне, а также статистика, связанная с Повесткой дня в области устойчивого развития на период до 2030 года. Общее описание подхода к анализу данных НЗ приводится в разделе 3.3. Этот подход предусматривает три обширных этапа: предварительную обработку изображений и данных, анализ данных для получения требуемых оценок и проверку результатов. Эти шаги более подробно описаны в разделах 3.4-3.7. Глава завершается кратким обобщением в разделе 3.8.

3.2 Данные НЗ и официальная статистика

Цели в области устойчивого развития:

Использование данных НЗ для целей официальной статистики соответствует Повестке дня в области устойчивого развития на период до 2030 года посредством использования более широкого круга данных в поддержку достижения целей, задач и показателей устойчивого развития (<http://unstats.un.org/sdgs/indicators/indicators-list/>). Как указано в разделе 2.19, данные НЗ могут способствовать достижению многих целей и показателей ЦУР. Примеры целей и показателей ЦУР, имеющих непосредственное отношение к данным НЗ, включают следующее:

- 2.4 К 2030 году обеспечить создание устойчивых систем производства продуктов питания и внедрить методы ведения сельского хозяйства, которые позволяют повысить жизнестойкость и продуктивность и увеличить объемы производства, способствуют сохранению экосистем, укрепляют способность адаптироваться к изменению климата, экстремальным погодным явлениям, засухам, наводнениям и другим бедствиям и постепенно улучшают качество земель и почв.
 - 2.4.1 Доля площади сельскохозяйственных угодий, на которых применяются продуктивные и неистощительные методы ведения сельского хозяйства.
- 6.3. К 2030 году повысить качество воды посредством уменьшения загрязнения, ликвидации сброса отходов и сведения к минимуму выбросов опасных химических веществ и материалов, сокращения вдвое доли неочищенных сточных вод и значительного увеличения масштабов рециркуляции и безопасного повторного использования сточных вод во всем мире.
 - 6.3.2 Доля водоемов с хорошим качеством воды.

- 6.6 К 2020 году обеспечить охрану и восстановление связанных с водой экосистем, в том числе гор, лесов, водно-болотных угодий, рек, водоносных слоев и озер.

- 6.6.1 Динамика изменения площади связанных с водой экосистем.

- 15.1 К 2020 году обеспечить сохранение, восстановление и рациональное использование наземных и внутренних пресноводных экосистем и их услуг, в том числе лесов, водно-болотных угодий, гор и засушливых земель, в соответствии с обязательствами, вытекающими из международных соглашений.

- 15.1.1 Площадь лесов в процентном отношении к общей площади суши.
 - 15.1.2 Доля важных с точки зрения биологического разнообразия районов суши и пресноводных районов, находящихся под охраной, в разбивке по видам экосистем.

- 15.2. К 2020 году содействовать внедрению методов рационального использования всех типов лесов, остановить обезлесение, восстановить деградировавшие леса и значительно расширить масштабы лесонасаждения и лесовосстановления во всем мире.

- 15.2.1 Прогресс в переходе на неистощительное ведение лесного хозяйства.
 - 15.2.2. Чистая безвозвратная потеря леса.

- 15.3 К 2030 году вести борьбу с опустыниванием, восстановить деградировавшие земли и почвы, включая земли, затронутые опустыниванием, засухами и наводнениями, и стремиться к тому, чтобы во всем мире не ухудшалось состояние земель.

- 15.3.1 Отношение площади деградировавших земель к общей площади земель

Anderson и др. (2017 г.) приводят ряд примеров, иллюстрирующих использование НЗ на службе Целей устойчивого развития на период до 2030 года. Выделены следующие преимущества: возможность принятия обоснованных решений и мониторинга ожидаемых результатов; улучшение национальной статистики для большей точности; гарантия того, что данные являются «пространственно-явными»; содействие синергии между ЦУР и многосторонними природоохранными соглашениями путем решения сквозных тем, таких как климат и энергетика; а также облегчение подходов стран к работе в различных секторах развития.

3.2.1 Статистика окружающей среды и сельского хозяйства

В настоящее время основное внимание при анализе данных НЗ уделяется получению статистических данных об окружающей среде и сельском хозяйстве. Три конкретных примера статистических данных, которые могут быть получены из данных НЗ – это данные о земном покрове, землепользовании и изменении земного покрова. Они определены в отчете Глобальной стратегии ФАО¹⁶ следующим образом: земной покров – это наблюдаемый (био)физический покров на поверхности Земли; землепользование – это использование земли людьми; а в ходе исследования изменений земель пытаются описать, где и какие изменения происходят с течением времени, с какой скоростью и, возможно, почему.

Другая сельскохозяйственная статистика включает идентификацию сельскохозяйственных культур и урожайность. Идентификация культур и тематические исследования урожайности подробно описаны в Главе 4. Для получения дополнительной информации об использовании данных НЗ с целью идентификации культур и урожайности см. Справочник ФАО по дистанционному зондированию для сельскохозяйственной статистики.

¹⁶ Глобальная стратегия 2017 г. Справочник по дистанционному зондированию для сельскохозяйственной статистики. www.gsars.org

Эти статистические данные можно свести в таблицу или отобразить графически, как показано на рис. 18 для Чьяпаса, Мексика.

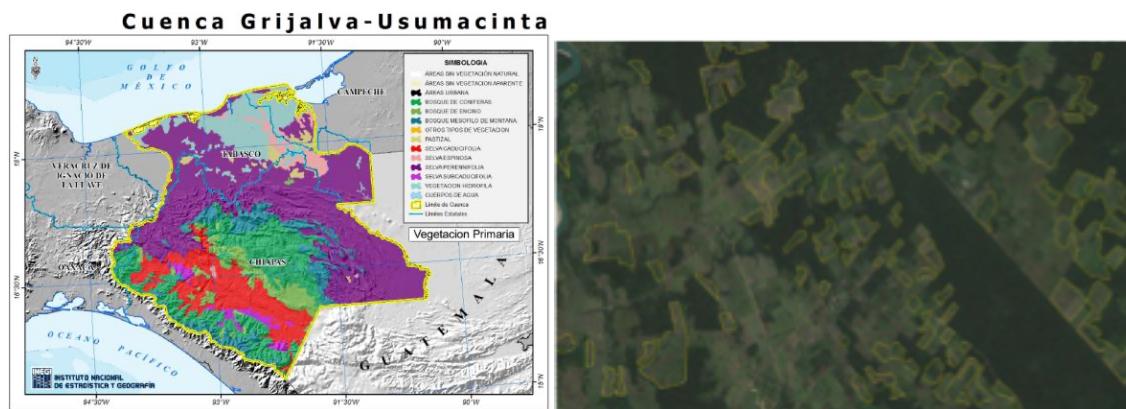


Рисунок 18: Примеры географического отображения статистических данных, относящихся к показателям ЦУР, которые могут быть получены полностью или частично на основе данных НЗ. Использование серии карт землепользования и растительности INEGI на цифровой карте Мексики (доступны различные масштабы). Слева: Показатель 15.2.1 Прогресс в переходе на неистощительное ведение лесного хозяйства. Справа: индикатор 15.2.2 Чистая безвозвратная потеря леса, оценка изменений в тропических широколиственных вечнозеленых лесах в штате Чьяпас на основе спутниковых снимков с 2006 по 2013 год.

Источник: Данные взяты из публикации «Интеграция статистической и геопространственной информации: опыт Мексики». Презентация г-на Rolando Ocampo, вице-президента INEGI, Мексика, в ходе организованного СОООН совещания по геопространственной информации и наблюдениям за Землей: поддержка официальной статистики в мониторинге ЦУР, Нью-Йорк, 7 марта 2016 г.

Еще одним примером графического отображения статистических данных об окружающей среде и сельском хозяйстве, полученных на основе спутниковых изображений, является карта глобального изменения лесов, составленная Hansen, Potapov, Moore, Hancher и др. (2013 г.). Потеря леса, потеря и прирост лесного покрова, а также процент древесного покрова определяются на основе анализа временных рядов изображений Landsat и отображаются на интерактивной глобальной карте в Google Earth Engine. Фрагмент карты показан ниже на рисунке 19.

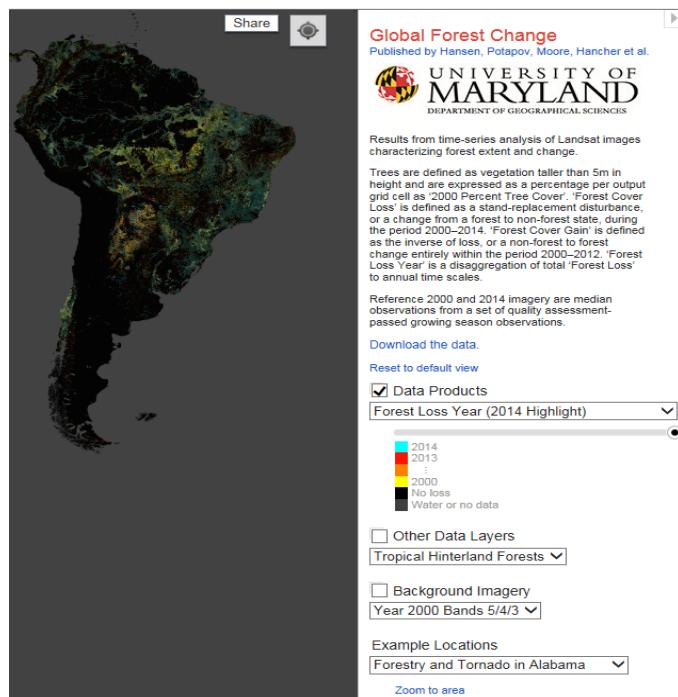


Рисунок 19: Карта глобального изменения лесов в Южной Америке

Для получения дополнительной информации и просмотра карты см. <https://earthenginepartners.appspot.com/science-2013-global-forest>.

3.2.2 Другая статистика

Многие другие статистические данные, представляющие интерес для НСУ, могут быть получены из данных НЗ. Примеры, касающиеся ЦУР, включают определение качества воды (Dekker и др., 2016 г.), городские изменения (Taubenboch и др., 2012 г.; Esch и др., 2012 г.) и оценку риска стихийных бедствий (Geib & Taubenboch, 2013 г.). Anderson и др. (2017 г.) также описывают примеры глобального мониторинга посевов (с помощью GeoGLAM), доступа к электричеству (визуализированного посредством использованияочных спутниковых изображений для оценки местоположения и распространения электрического освещения) для ЦУР 7.1, качества воздуха (путем мониторинга аэрозолей, горячих точек и лесных пожаров), ЦУР 11.6) и биоразнообразия (посредством использования НЗ для мониторинга охраняемых территорий и потенциальной преступной деятельности, ЦУР 14 и 15).

Различные формы данных НЗ могут быть объединены для получения соответствующей статистики. Например, в настоящее время беспилотные летательные аппараты широко используются в качестве источников данных НЗ сами по себе и в сочетании со спутниковыми и радиолокационными данными. Они также используются в сочетании с традиционными источниками данных, такими как опросы, а также с другими источниками больших данных, такими как данные мобильных телефонов и социальных сетей. Например, Steele и др., (2017 г.) приводят пример объединения данных НЗ и мобильных телефонов для получения оценок бедности.

Важно отметить, что различные определения интересующих величин могут повлиять на расчет оценок. Это описано в отчете Глобальной стратегии ФАО в отношении изменения земного покрова: существенные различия были обнаружены в 11 классах покрова, полученных из четырех различных глобальных баз данных. В отчете отмечается, что «это не означает, что цифры неточны (поскольку они относятся к разным определениям принадлежности к классу); скорее, они непоследовательны, особенно там, где используется мало классов». По этой причине ФАО и UNEP (Программа ООН по окружающей среде) разработали общую систему классификации покрова земли (LCCS)¹⁷, как подробно описано в отчете¹⁸. Центральная основа СЭЭУ¹⁹ является полезным международным стандартом экологического учета, и ее классификация земного покрова основана на наборе правил и классификаторов, изложенных в LCCS.

3.3 Общий подход к анализу данных НЗ

3.3.1 Этапы анализа данных НЗ

Анализ данных НЗ можно разделить на три основных этапа:



Этап **предварительной обработки** имеет решающее значение для успешного анализа больших данных, а также для качества и достоверности полученных оценок. Этот шаг включает в себя следующие соображения:

¹⁷ См. также LCCS2 <http://www.fao.org/docrep/008/y7220e/y7220e00.htm> и LCCS3 <http://www.fao.org/3/a-i5232e.pdf>

¹⁸ Ссылка на отчет: http://gsars.org/wp-content/uploads/2016/08/TR_Information-on-Land-in-the-Context-of-Ag-Statistics-180816.pdf

¹⁹ Центральная основа СЭЭУ: https://unstats.un.org/unsd/envaccounting/searev/seea_cf_final_en.pdf

- *Хранение данных:* изображений НЗ и/или извлеченных данных, а также статистических данных и производных продуктов, полученных в результате анализа данных.
- *Управление данными:* изображениями НЗ и извлеченными данными.
- *Качество данных:* критическая оценка данных, извлеченных из изображений НЗ, в отношении их общей способности к анализу и способности предоставлять надежные и точные статистические данные.
- *Другие источники информации:* идентификация, сопоставление, хранение, управление и оценка других вспомогательных данных, которые можно использовать на этапе анализа.
- *Программное обеспечение:* исходя из приведенных выше соображений, это включает в себя идентификацию подходящих доступных программных продуктов, которые можно использовать для помощи на этапе предварительной обработки.

Этап **анализа** включает в себя следующие соображения:

- *Общая цель:* определение общей цели, такой как предоставление конкретных национальных статистических данных или показателя ЦУР; и определение соответствующих статистических оценок, прогнозов или выводов, которые должны быть получены в результате анализа.
- *Данные:* определение того, какое подмножество сохраненных данных будет использоваться в анализе, и будет ли анализ основан исключительно на данных НЗ или в сочетании с другими источниками данных.
- *Аналитический метод:* выбор общего подхода и конкретной техники для извлечения требуемых количеств, исходя из цели и имеющихся данных.

Этап **оценки** включает в себя следующие соображения:

- *Критическая оценка результатов:* сопоставление, оценка и интерпретация результатов анализа.
- *Оценка точности:* сравнение наблюдаемых значений и прогнозируемых оценок, где это возможно, оценка неопределенности или точности результатов и оценка воспроизводимости результатов.
- *Другие опасения:* рассмотрение любых других потенциальных предубеждений или опасений, связанных с результатами и их интерпретацией.

Эти шаги подробно описаны в разделах с 3.4 по 3.7.

3.3.2 Внедрение методологии

Различные национальные статистические службы разработали планы проведения анализа данных НЗ. Эти планы включают вышеуказанные шаги по-разному.

В качестве иллюстрации на Рисунке 20 показан шаблон, предложенный Данией для облегчения перехода от сбора данных НЗ к оценке соответствующих статистических данных и интеграции в операции НСУ.

Предлагаемая интеграция геопространственных данных
Анализ пробелов: «До сих пор отсутствуют надежные методы оценки выбросов в результате деградации лесов. Мы предлагаем использовать общий источник ввода (Sentinel 2) для наблюдения за мировым лесом, что устраниет необходимость в многоуровневой системе. Это также стандартизирует методы отчетности и повысит прозрачность, опираясь на тесное сотрудничество с национальными органами управления лесным хозяйством».
Перечислите необходимые геопространственные данные: «Потребность в мультиспектральных изображениях высокого разрешения (включая NIR) для подробных изображений земли и растительности с частым повторным посещением для получения частых изображений».
Требования к качеству данных: «Этот индикатор требует высокой частоты повторения для получения большого охвата данных за короткие периоды времени (короткий цикл повторения), высокого пространственного разрешения (10-20 м) для оценки лесонасаждений с низким уровнем сомнительности лесного полога, 10, 20 и 60 м, и высокое спектральное разрешение для различия лесов и спектрально схожих типов растительности».
Доступность данных: «Данные Sentinel доступны по всему миру, их можно загрузить с портала ESA. Доступ к данным Sentinel является бесплатным, полным и открытым для широкого регионального, национального, европейского и международного сообщества пользователей. Регистрация пользователя основана на предварительной регистрации учетной записи пользователя с отдельной учетной записью для каждого Соглашения».
Сбор данных: «Инфраструктура доступа к данным Sentinel для международных соглашений (Хаб данных по международным соглашениям) может предоставить доступ к скользящему онлайн-архиву, охватывающему последние месяцы основных продуктов Sentinel, доступных с определенной своевременностью. Кроме того, по запросу предоставляется доступ к архивным данным в автономном режиме».
Интерпретация данных: «Процедура оценки изменения лесного покрова: получение данных НЗ, контроль и предварительная обработка изображения участка, предварительная маркировка объектов и изменений, проверка и корректировка маркировки, проверка и добавление параметров лесопользования и землепользования».
Метод интеграции: «1) Структура управления согласовывается на национальном и международном уровне, 2) Создается глобальный набор справочных данных, 3) Согласуются циклы мониторинга, 4) Разрабатываются методы обнаружения изменений, и пересматривается центрально установленный набор данных, 5) Интернет-портал, подобный Системе управления информацией о лесных ресурсах «FRIMs», используется в качестве канала взаимодействия между ФАО и каждым национальным органом».

Рисунок 20: Шаблон предложен Агентством по предоставлению данных и эффективности в Дании.

Источник: Презентация г-на Morten Nordahl Moller, начальника отдела, в ходе организованного СОООН совещания по геопространственной информации и наблюдениям за Землей: поддержка официальной статистики в мониторинге ЦУР, Нью-Йорк, 7 марта 2016 г.

3.4 ЭТАП 1: Предварительная обработка

Соображения, которые необходимо рассмотреть на этапе предварительной обработки, включают (1) хранение данных НЗ, (2) управление данными, (3) оценку качества данных, (4) идентификацию других источников информации и (5) выбор программных продуктов, помогающих на этом этапе. Теперь рассмотрим эти соображения подробно.

3.4.1 Хранение данных НЗ

Традиционно НСУ собирали выборочные данные для получения статистических результатов и выводов о более широких группах населения. Источники больших данных, такие как данные НЗ, открывают возможности для более своевременного доступа к большим объемам данных и часто с меньшими затратами. Однако наборы данных НЗ велики, и поэтому необходимы соответствующие подходы к их хранению. Рекомендации по решениям и подходам к хранению выходят за рамки данного отчета, поэтому здесь кратко упоминаются только несколько вариантов.

Варианты хранения для НСУ включают:

- Локальное хранение данных;
- Аутсорсинг хранения данных.

Для локального хранения данных обычно требуется какая-либо среда высокопроизводительных вычислений (HPC). Кластеры высокопроизводительных вычислений и методы распределенных

вычислений позволяют разбивать большие наборы данных на блоки и запускать их одновременно на нескольких узлах для сокращения времени вычислений. Затем результаты каждого узла объединяются статистически обоснованным образом.

Настройка системы высокопроизводительных вычислений, способной работать с текущими и будущими данными НЗ, может быть дорогостоящей с точки зрения инфраструктуры, обучения, и обслуживания. По этим причинам аутсорсинг хранения данных является привлекательной альтернативой. Например, существуют различные облачные решения для хранения данных, предлагаемые Microsoft и Amazon Web Services, или можно рассмотреть местные проверенные компании.

Другой вариант – прибегать к помощи третьих лиц или использовать открытые данные. Открытые данные НЗ становятся все более распространенными и в настоящее время входят в компетенцию многих источников данных НЗ (например, Copernicus). Использование готовых к анализу данных может еще больше упростить процесс, устранив не только проблемы хранения, но и большую часть нагрузки по предварительной обработке. Кубы данных, описанные в главе 2, например, обладают возможностями НРС и делают стандартизованные и откалиброванные спутниковые данные доступными в структурах, на различных платформах обработки, которые могут быть локальными, удаленными или облачными. Онлайновые и бесплатные каталоги данных в сочетании с онлайновыми платформами моделирования, такими как GEE – Code Editor, также могут быть применимы.

3.4.2 Управление данными НЗ

Управление данными НЗ обычно включает в себя обработку изображений для готовых к анализу данных (ARD).

Как описано в главе 2 и в предыдущем разделе, в некоторых странах доступны кубы данных или другие источники ARD. Обычно это можно проанализировать без дальнейшей предварительной обработки. Однако важно, чтобы другие соображения, такие как оценка систематической ошибки и проверка результатов, по-прежнему принимались во внимание. Эти вопросы более подробно описаны в последующих разделах этой главы.

Для многих НСУ данные ARD могут быть еще недоступны. Как обсуждалось в главе 2, исходные данные НЗ нельзя использовать для анализа в необработанном виде, поскольку это привело бы к неточным результатам, основанным на различных углах солнца и геометрии датчиков, шуме, дымке и искажении исходных данных изображений. Для повышения качества получаемых изображений проводятся различные уровни предварительной обработки.

Цель предварительной обработки состоит в извлечении чистой спектральной информации, отраженной с поверхности Земли и записанной датчиком без каких-либо скрытых аспектов. Это делается с помощью так называемых алгоритмов коррекции, а именно геометрической коррекции и радиометрической коррекции. Эти алгоритмы осуществляют фильтрацию и коррекцию для конкретных и определенных параметров шума. Цель геометрической коррекции заключается в повышении точности поверхностного спектрального отражения, эмиссии или обратного рассеяния измерений с использованием системы дистанционного зондирования. Этот этап предварительной обработки также известен как исправление ошибок детектора или атмосферная и топографическая коррекция. Радиометрическая коррекция касается атмосферных частиц, которые могут скрывать спектральную информацию.

В дополнение к этим поправкам, внесенным в спутниковые снимки с верхних слоев атмосферы, требуется также «маскировка» данных. Маскировка относится к процессу обнаружения, а затем «маскированию» затеняющих артефактов, таких как облачный покров, тень облаков, пирокумулятивные облака, вода и так далее. Все эти исправления являются неотъемлемой частью создания готовых к анализу данных (ARD).

Потратив немного времени и навыков, национальные статистические службы могут настроить систему для обработки исходных данных спутниковых изображений на этапе ARD с использованием либо патентованного, либо программного обеспечения с открытым исходным кодом.

Программное обеспечение с открытым исходным кодом потенциально дешевле в установке и обладает тем преимуществом, что обеспечивает независимость от несвободных систем. Это дает преимущества, позволяя настраивать программное обеспечение и увязывать его с более широким кругом программных продуктов и систем, которые уже внедрены в организации. Кроме того, часто существует большое сообщество разработчиков, работающих с открытым исходным кодом, которое может предоставить поддержку и варианты решения проблем. С другой стороны, платные программы могут быть более надежными с точки зрения клиента или потребителя и могут иметь преимущество более профессиональной технической поддержки (Trevithick & Gillingham, 2009 г.). Программные продукты рассматриваются далее в разделе 3.4.5.

Подробная система, которая использует программное обеспечение с открытым исходным кодом для обработки исходных данных спутниковых изображений в ARD, описана в разделе 3.4.6. Эта система была разработана группой дистанционного зондирования Департамента науки, информационных технологий и инноваций Квинсленда (DSITI) в Австралии. Система была разработана до создания Австралийского куба данных и активно используется уже более десяти лет. Она использует данные Landsat, доступные на веб-сайте Геологической службы США (USGS), и применяет ряд пакетов программного обеспечения с открытым исходным кодом, которые переводят форматы данных, вносят атмосферные поправки и программируют обработку изображений. Пример того, как эти обработанные данные используются для определения сельскохозяйственной деятельности и землепользования, представлен в качестве тематического исследования в Главе 4.

3.4.3 Оценка качества данных НЗ

Традиционно НСУ собирали выборочные данные для получения статистических результатов и выводов о более широких группах населения. Источники больших данных, такие как данные НЗ, открывают возможности для более своевременного доступа к большим объемам данных и часто с меньшими затратами.

Существует заблуждение, что источники больших данных, такие как данные НЗ, будут статистически достоверными просто из-за объема доступных данных. Хотя увеличение размера выборки может уменьшить ошибки выборки, оно не уменьшает другие источники статистической погрешности, такие как ошибка измерения. Кроме того, наборы данных, полученные из этих источников, не обязательно являются случайными выборками из целевой совокупности, о которой НСУ хотят сделать выводы, поэтому увеличение размера набора данных может не только не уменьшить ошибку выборки, но и не улучшить качество оценок.

Кроме того, у источников больших данных есть свои специфические проблемы, которые могут повлиять на статистическую достоверность. К ним относятся накопление ошибок (шум) и потенциальные ложные корреляции, а также алгоритмическая нестабильность, которая может привести к неточным оценкам. Различные источники данных и их агрегирование также могут привести к статистическим ошибкам (Fan, Han & Liu, 2014 г.).

Некоторые из этих проблем, связанных с использованием больших данных в статистических целях, можно решить с помощью соответствующих методологий в зависимости от источника данных и предполагаемых результатов. В этой главе описываются примеры различных методологических приемов для данных НЗ. Однако следует отметить, что эти методологии не будут решать проблемы, связанные с качеством данных, репрезентативностью данных или другие соответствующие проблемы. Это обсуждается далее в разделе 3.7.3.

3.4.4 Другие источники информации

Данные НЗ могут быть проанализированы сами по себе или с другими наборами данных. Выбор других наборов данных зависит от следующих соображений:

- *Цель*: для разных целей потребуются разные наборы данных, основанные на требуемых оценках, выбранном методе анализа, пространственном и временном разрешении, а также масштабе анализа, качестве доступных данных НЗ и т. д.
- *Доступ*: наборы данных могут включать существующие ресурсы, которыми владеет НСУ, и/или они могут быть получены от внешних источников. Вопросы доступа также включают интеллектуальную собственность, конфиденциальность, стоимость, репутацию и непрерывность доступа.
- *Содержание*: Широкий спектр аналитических данных потенциально может дополнять данные НЗ, например, переписи населения, обследования домашних хозяйств или сельскохозяйственных исследований, административные данные, экологические и метеорологические данные.

В отчете ФАО о Глобальной стратегии совершенствования сельскохозяйственной и сельской статистики (2016 г.) содержатся рекомендации по созданию и ведению минимального набора основных данных (MSCD), включающего данные НЗ и другую пространственную и непространственную информацию в основе базы данных эталонной выборки (MSF). MSF строится путем объединения спутниковых изображений, классифицированных по землепользованию, и оцифрованных административных данных, оцифрованных переписных участков и наложения их на данные переписи населения и сельскохозяйственной переписи, а также других соответствующих данных, таких как реестр коммерческих ферм.

(http://gsars.org/wp-content/uploads/2016/08/TR_Information-on-Land-in-the-Context-of-Ag-Statistics-180816.pdf, Информация о земле в контексте сельскохозяйственной статистики, Серия технических отчетов GO-15-2016)

Многие НСУ находятся на различных этапах внедрения систем для интеграции данных НЗ с другими данными для официальной статистики и ЦУР. Например, в Таблице 5 показано, как Мексика планирует использовать спутниковые изображения разного разрешения для целей национальной статистики и для поддержки соответствующих ЦУР. Эти статистические данные можно отображать географически, как показано на рисунке 18.

Таблица 5: Объединение статистической и геопространственной информации: опыт Мексики.

ГЕОСТАТИСТИЧЕСКАЯ ИНТЕГРАЦИЯ для МОНИТОРИГА НАЦИОНАЛЬНЫХ ПРИОРИТЕТОВ И ЦУР			
Спутниковые изображения	Другие источники	Национальное использование	ЦУР / другие применения
Высокое разрешение (2,5 м) SPOT ERMEX Очень высокое разрешение (0,5 м) GEOEYE EVISMAR	Перепись населения	Геостатистическая основа Обновление: Топографических диаграмм	1. Ликвидация нищеты 2. Ликвидация голода 3. Хорошее здоровье и благополучие 4. Качественное образование 5. Гендерное равенство 6. Чистая вода и санитария 7. Недорогостоящая и чистая энергия 8. Достойная работа и экономический рост 9. Индустриализация, инновации и инфраструктура 10. Уменьшение неравенства 11. Устойчивые города и населенные пункты 12. Сохранение экосистем суши
	Национальная жилищная инвентаризация	Визуализация: Населения, Экономики, Домохозяйств, Гендер, Здравоохранения, Образования и общественных услуг	
	Экономическая перепись	Инфраструктура водоснабжения и санитарных услуг	
	Технический стандарт по адресам	Выбросы ПГ/опасных отходов	
	COA-Web		
	Валидация на месте		
Среднее Разрешение (5-30 м)	Природные ресурсы и топографические диаграммы	Серия карт по Землепользованию и Растительному покрову Обезлесение, изменения землепользования	2. Ликвидация голода 6. Чистая вода и санитария 13. Борьба с изменением климата 14. Сохранение морских экосистем 15. Сохранение экосистем суши
RAPIDEYE LANDSAT	Данные о лесном и водном секторах		

	Валидация на месте	Мониторинг с/х культур	
Низкое разрешение (250 м) MODIS	Топографические карты Землепользование и растительный покров	Мониторинг бедствий Пожары, наводнения	Сендайская рамочная программа Борьба с изменением климата
Радар RADARSAT		Мониторинг бедствий Наводнения, цифровые модели в туманных районах	Сендайская рамочная программа Борьба с изменением климата

Источник: Презентация г-на Rolando Osampo, вице-президента INEGI, Мексика, в ходе организованного СОООН совещания по геопространственной информации и наблюдениям за Землей: поддержка официальной статистики в мониторинге ЦУР, Нью-Йорк, 7 марта 2016 г.

3.4.5 Программные продукты

Существует ряд запатентованных и открытых программных продуктов, доступных для работы с данными НЗ. Продукты могут использоваться для помощи в подготовке, хранении и управлении данными. В настоящей главе не содержится рекомендаций по программным продуктам, поскольку наиболее подходящий инструмент будет различным для отдельных НСУ.

При выборе продукта НСУ должны учитывать возможности и опыт своего персонала, использующего различное программное обеспечение, стоимость при выборе коммерческого продукта и долгосрочную доступность при выборе продукта с открытым исходным кодом. Например, если НСУ собирает статистические данные на основе данных, полученных с использованием продукта с открытым исходным кодом, то как на них повлияет, если провайдер начнет взимать плату за доступ к данным, и есть ли у НСУ планы на случай непредвиденных обстоятельств.

Список доступного программного обеспечения быстро увеличивается. Некоторые популярные продукты, доступные сейчас, кратко описаны здесь.

Hadoop – это популярная технология управления большими данными (Scannapieco, Virgillito & Zardetto (2013 г.). Это проект с открытым исходным кодом, разработанный Apache Software Foundation, направленный на поддержку выполнения приложений, ориентированных на данные, на кластерах общего оборудования. Основными компонентами Hadoop являются файловая система HDFS (распределенная файловая система Hadoop), которая отправляет задачи на несколько ядер для обработки, и MapReduce, которая отслеживает пакеты данных и объединяет результаты процессов. В настоящее время доступно множество продуктов, использующих Hadoop в качестве платформы и предоставляющих улучшенные интерфейсы для облегчения взаимодействия с конечными пользователями. Популярные текущие примеры включают Spark и его производные (такие как Sparkling Water и Steam) и Tensor Flow.

GDAL/OGR – Библиотека абстракции геопространственных данных (чтение и запись растровых и векторных форматов геопространственных данных) – большинство широко используемых геопространственных приложений используют GDAL/OGR.

GDAL поддерживает более 140 растровых форматов; OGR поддерживает более 80 векторных форматов (см. дополнительную информацию в публикации Warmerdam и др., 2016 г.).

Некоторые программные пакеты (например, растр в R (R Core Team, 2016 г.)) позволяют манипулировать данными, слишком большими, чтобы загружать их в память в полном объеме, только создавая структурную информацию о данных, и обрабатывать фрагменты при работе с самими данными.

GIS – система автоматизированного геолого-научного анализа (SAGA) SAGA – это программа с открытым исходным кодом, предназначенная для предоставления ученым эффективной, но легко обучаемой платформы для реализации геолого-геофизических методов и позволяющей пользователю выполнять векторный и растровый пространственный анализ, редактировать и

манипулировать разнообразными данными и хранить его в формате, позволяющем делиться им с другими GIS-сообществами и коммерческими программными пакетами. Кроме того, она может выполнять интерполяции с использованием триангуляции, ближайшего соседа, обратного расстояния и других метрик. SAGA GIS можно загрузить по ссылке <http://www.sagagis.org/en/index.html>.

SAGA имеет быстрорастущий набор геолого-геофизических методов и написана на C++. Она была разработана на кафедре физической географии Геттингенского университета в Германии и в настоящее время поддерживается и расширяется международным сообществом разработчиков. Первый релиз программы состоялся в феврале 2004 года.

Google Earth Engine

Как описано в разделе 3.2.1, Google Earth Engine можно использовать для отображения и интерпретации данных НЗ. Данные Landsat теперь доступны в Earth Engine благодаря сотрудничеству с Геологической службой США (USGS) и NASA. Earth Engine предоставляет эти данные в исходном виде, как отражательную способность с поправкой на TOA, а также в различных продуктах, рассчитанных с помощью ARD, таких как вегетационные индексы NDVI и EVI. Более подробно см. <https://earthengine.google.com/datasets>.

Расширения существующих программных продуктов

Растущее число программных продуктов, которые традиционно имели дело только с базами данных, теперь предоставляют возможности работы с большими данными и инструменты управления данными. Примеры включают SAP (Hana) и HPE (Vertica). Точно так же пакеты статистического и математического моделирования, такие как R, Matlab, Stata и SAS, теперь предоставляют инструменты управления большими данными, часто основанные на Hadoop и аналогичных платформах.

Также можно найти специализированные инструменты управления данными и/или написанные на таких языках программирования, как R (см., например, RStools pack) и Python.

Дополнительные примеры программного обеспечения, используемого для предварительной обработки данных НЗ, обсуждаются в тематическом исследовании, описанном в следующем разделе.

3.4.6 Практический пример: предварительная обработка данных

В Австралии Департамент науки, информационных технологий и инноваций (DSITI) Квинсленда создал систему, использующую программное обеспечение с открытым исходным кодом для обработки первичных данных спутниковых изображений в ARD. Это было еще до создания Австралийского куба данных.

Пакеты программного обеспечения, которые имеют решающее значение для процесса – это GDAL, 6S и Python. Это программы с открытым исходным кодом. GDAL – представляет собой библиотеку-транслятор для растровых и векторных форматов геопространственных данных. 6S – это программное обеспечение для моделирования переноса излучения в атмосфере, которое используется для внесения атмосферных поправок. Python – язык программирования, используемый для большей части обработки изображений в этом исследовании.

Во-первых, исходные данные Landsat получены для всего штата Квинсленд с веб-сайта Геологической службы США (USGS). Данные Landsat включаются, если облачность <80%. Затем данные в файлах распаковываются в отражающие, термические и панхроматические компоненты и сохраняются в виде файлов ERDAS Imagine (.img). Также можно использовать другие форматы, читаемые программным обеспечением GDAL. Обработка тепловых, панхроматических и отражающих компонентов кратко описана ниже.

Термический компонент

Термическая составляющая настраивается на температуру, а маски облаков и теней применяются автоматически. Затем автоматически применяются маски снега, облаков и теней от облаков. Затем эти маскированные данные включаются в автоматический этап частичного сезонного покрытия в процессе отражающего компонента. Это входные данные для облаков и масок теней облаков, которые применяются для создания ARD.

Панхроматический компонент

Это одноэтапный процесс, в котором исходные данные обрабатываются в Python для создания панхроматического изображения. Этот компонент далее в анализе не используется.

Отражающий компонент

Из необработанных данных Landsat автоматически создается масштабированное изображение яркости верхней части атмосферы. Затем это изображение яркости атмосферы корректируется с учетом атмосферных эффектов и сохраняется как img-файл.

Согласно Flood и др. (2013 г.), «атмосферное воздействие на сигнал включает в себя ряд процессов рассеяния и поглощения и зависит от различных углов, связанных с солнцем, поверхностью и спутником. Предполагается, что эти эффекты хорошо моделируются широкодоступными моделями переноса излучения в атмосфере».

В работе DSITI используется модель второго моделирования спутникового сигнала в солнечном спектре (широко известная как 6S) для применения атмосферных поправок и получения скорректированного img-файла. Дополнительные сведения о применении модели 6S см. у Flood и др. (2013 г.), раздел 2.2.

Скорректированный img-файл с различными примененными масками облачного покрова, тени от облаков, падения солнечных лучей, топотени и воды является результирующим ARD, используемым в анализе.

После завершения этих процессов данные готовы для анализа. Обратите внимание, что в ARD включен только отражающий компонент, а в него включено только подмножество отражающих изображений с облачным покровом <60%.

3.5 ЭТАП 2: Анализ

Прежде чем приступить к анализу, следует рассмотреть следующие вопросы:

1. Какие требуется меры?
2. Какие статистические данные будут использоваться для оценки этих показателей?
3. Какие данные необходимы для получения этой статистики?

Первые два вопроса были рассмотрены выше в контексте НСУ и ЦУР; см., например, раздел 3.2. При решении третьего вопроса важно признать, что не все имеющиеся данные требуются для конкретного анализа. Хорошая структура управления данными позволит либо агрегировать соответствующие данные для анализа, либо, в качестве альтернативы, предоставить инструменты для доступа к наборам данных из разных баз данных, хранящихся в разных (виртуальных) местах. У обоих этих подходов есть преимущества и недостатки. Идентификация только релевантных данных для анализа помогает защититься от получения ложных результатов из-за включения лишней информации.

3.5.1 Аналитический подход

Помимо определения типа аналитической задачи, необходимо выбрать аналитический подход (или комбинацию подходов). Как обсуждалось в разделе 2.4, аналитические подходы можно разделить на пять типов методов:



Рисунок 21: Аналитические подходы к анализу НЗ и других данных.

Выбор аналитического подхода зависит от имеющихся данных и степени понимания физических процессов, лежащих в основе взаимосвязи между входными данными НЗ и целевыми оценками. Например, если есть четкое понимание биологических или физических процессов, то основанный на физике подход является пригодным, в противном случае может быть принят эмпирический подход или подход машинного обучения. Если данные доступны и есть знания о процессе, то можно использовать полуэмпирические методы.

Хотя пять методологических подходов имеют различные характеристики, они существенно пересекаются. Например, хотя эмпирические подходы больше сосредоточены на моделях, а подходы машинного обучения – на алгоритме, оба подхода направлены на создание взаимосвязей на основе наблюдаемых данных, и для этого они используют общий набор методов. Точно так же математические и статистические модели, как правило, рассматривались как разные, но теперь принято объединять математические или физические методы со статистическими моделями и моделированием, где это уместно. Например, математическое представление процесса может сочетаться со статистическими описаниями менее известных частей процесса и/или неопределенностью входных параметров. Затем может быть включено моделирование для описания более сложных компонентов или дальнейшей оценки неопределенности входных и выходных данных или оценки сценариев «что, если».

Эмпирические методы

Эмпирические методы определены в Главе 2 как случаи, когда устанавливается статистическая связь между используемыми спектральными каналами или частотами и измеряемой переменной (на основе поля), при этом не обязательно наличие причинно-следственной связи. Эта связь может быть параметрической, полупараметрической или непараметрической.

Основные преимущества эмпирических подходов (основанных на статистических моделях) заключаются в том, что они обеспечивают математически строгий способ описания ошибок выборки и модели, оценки и прогнозирования интересующих результатов и взаимосвязей между переменными, а также количественной оценки неопределенности, связанной с этими оценками и прогнозами. Их также можно использовать для проверки гипотез при определенных предположениях. Недостатки заключаются в том, что они часто включают предположения о процессе генерации данных (в отличие от многих методов машинного обучения) и требуют

локальных данных (в отличие от моделей, основанных на физике). Эта зависимость от данных в месте нахождения также означает, что их может быть трудно экстраполировать или перенести в другие контексты (Dekker и др., 2013 г.).

Помимо пространственного охвата, многие формы данных дистанционного зондирования (особенно со спутников) собираются с течением времени. Хотя добавление временного измерения существенно увеличивает размер данных, с этим можно справиться путем тщательного отбора данных (например, на основе *aприорных* знаний о конкретных циклах роста сельскохозяйственных культур). Включение временного измерения в анализ может добавить множество полезных сведений, таких как существенно более точная классификация сельскохозяйственных культур, а также оценка и прогнозирование урожайности.

Полуэмпирические методы

Полуэмпирические методы сочетают знания о процессе со статистическими моделями. Это знание может быть о самом процессе или о переменных, которые включены в процесс.

Эти методы использовались для анализа данных НЗ более десяти лет. Например, Phinn и др. (2005 г.) использовали данные изображений Landsat 7 Enhanced Thematic Mapper для картирования выбранных параметров качества воды и типа субстрата. Совсем недавно Dekker и др. (2013 г.) описывали полуэмпирический подход к определению качества воды, при котором для уточнения статистической модели используются знания о спектральных характеристиках некоторых параметров. В их примере в модели используется только подмножество переменных с хорошо выбранными спектральными областями и соответствующими диапазонами волн или комбинациями диапазонов волн. Авторы подчеркивают популярность и полезность этих подходов, но также предупреждают, что они по-прежнему требуют локальных данных. Таким образом, хотя они, возможно, более применимы, чем чисто эмпирические модели (поскольку содержат информацию о процессе), они все же могут быть ограничены общностью данных, используемых для их построения.

Tripathy и др. (2014 г.) также используют полуэмпирический метод, который включает физиологические показатели, спектральные показатели и пространственные характеристики для оценки урожайности пшеницы. Авторы отмечают, что, хотя спектральные (эмпирические) и физические (механистические) модели, основанные на вегетативных индексах, широко используются, они, соответственно, ограничены, поскольку требуют больших объемов данных и сложны. Их полуэмпирический подход предлагается в качестве промежуточного метода.

Методы, основанные на физике

Как указано в главе 2, физические модели основаны на детальном знании моделируемой системы. Они могут быть построены без данных или данные могут использоваться для калибровки параметров модели. Этот метод подходит для автоматизации на больших площадях при условии, что модель соответствующим образом и точно параметризована.

Биофизические и геофизические модели, использующие данные НЗ, разрабатывались и применялись для решения широкого круга задач более 15 лет. Одним из первых примеров является расширение модели роста растений PHYSGROW для включения продуктов спутниковых данных NOAA и NASA для создания крупномасштабных карт производства кормов. В качестве исходных данных НЗ использовались данные о суточной температуре с привязкой к квадратам сетки, осадках и Нормализованном разностном вегетационном индексе (NDVI), а также данные о кокригинге с использованием пространственной автокорреляции (Angerer и др., 2001 г.). Авторы пришли к выводу о том, что картирование поверхностей на основе выходных данных кокригинга может позволить успешно идентифицировать районы засухи, и заявили, что эти карты могут использоваться в рамках географической информационной системы (GIS), которая затем может быть увязана с экономическими моделями, оценками рационального использования природных ресурсов или системами раннего предупреждения о засухе».

Phinn и др. (2005 г.) впервые применили эти методы в морском контексте, используя данные изображений Landsat 7 Enhanced Thematic Mapper для картирования выбранных параметров качества воды и типа субстрата. Более новый основанный на физике метод определения качества воды с использованием данных НЗ описан Dekker и др. (2013 г.). В настоящее время существует множество примеров таких моделей. Например, Watts и др. (2014 г.) расширили наземную модель потоков, которая позволяет использовать спутниковые данные в качестве основных входных данных для оценки потоков CO₂ и CH₄, а Gow и др. (2016 г.) объединили спутниковые наблюдения за температурой поверхности земли с моделью поверхностного энергетического баланса для оценки использования растительностью подземных вод.

Эти основанные на физике подходы, как правило, основаны на сложных алгоритмах, которые учитывают характеристики датчиков, многочисленные воздействия на окружающую среду со стороны атмосферы и поверхности моря, а также оптические свойства водоема и морского дна. Для получения дополнительной информации см. Wettle и др. (2013 г.).

Объектный анализ изображений

Два типа анализа изображений на основе объектов – это подходы агрегации на основе полей и подходы контекстной классификации.

Самый простой подход – агрегация на уровне полей. Этот метод объединяет пиксели до «полей» (классификация по полям) или до «объектов» (объектно-ориентированная классификация), а затем оценивает или классифицирует на этом новом агрегированном уровне. Эти подходы пространственно сглаживают данные путем усреднения по всем пикселям в поле или объекте. Это эффективно с точки зрения вычислений, а также, как было показано, дает более точные и надежные выходные данные, чем подходы на основе пикселей, особенно с пространственными данными с высоким разрешением. Например, Ноуэ и др. (2016 г.) и Ноуэ и др. (2017 г.) используют полевые подходы для моделирования воздействия циклонов, а в последнем документе оцениваются риски циклонов для существующих и будущих сценариев изменения климата.

Контекстные методы явно признают и включают пространственные зависимости между соседними пикселями. Они могут использовать пространственные сглаживатели, Марковские случайные поля, пространственную статистику и модели в сочетании с методами классификации. Оценка или классификация происходит либо после учета пространственных зависимостей (предварительное сглаживание), либо до (последующее сглаживание).

Методы искусственного интеллекта и машинного обучения

Методы машинного обучения, также называемые методами искусственного интеллекта (ML), используют алгоритмические модели для анализа данных. Эти методы рассматривают способ получения данных или отношения между переменными как неизвестные (Breiman, 2001 г.).

Существует множество алгоритмов машинного обучения, которые выполняют разные задачи. Некоторые из алгоритмов, которые имеют отношение к данным НЗ, примененным к ЦУР, перечислены на следующем рисунке и сгруппированы в соответствии с четырьмя основными аналитическими целями: классификация, группирование, регрессия и уменьшение размерности. Эти цели описаны в разделах 3.5.2 и 3.5.3. Подробнее об алгоритмах можно ознакомиться в Приложении.



Рисунок 22: Эмпирические методы и методы машинного обучения для анализа НЗ и других данных, сопоставленных в соответствии с аналитической целью.

Комплексные подходы

Последние тенденции в анализе машинного обучения и дистанционного зондирования сосредоточены на сочетании нескольких методов машинного обучения и эмпирических методов для формирования гибридных подходов. Эти комплексные подходы также известны под другими названиями, такими как множественные системы классификаторов для задач классификации (Du и др., 2012 г.; Woźniak и др., 2014 г.). Методы комплексного подхода делятся на две категории: последовательные (или конкатенация) и параллельные.

Последовательное комбинирование относится к последовательному комбинированию методов – результаты первого анализа используются в качестве входных данных для следующего анализа и так далее. Конечный результат (классификация, оценка, кластеризация, уменьшение размерности и т. д.) определяется выходными данными последнего метода в серии.

Параллельное комбинирование относится к подходу, который одновременно применяет несколько методов к данным; с окончательным выходом, определяемым с использованием некоторого решающего правила. Самым популярным правилом принятия решений для непрерывных результатов является простое среднее всех методов. Для систем с несколькими классификаторами наиболее популярным правилом является правило большинства, при котором окончательное членство в классе определяется большинством классификаторов. Выходные данные методов также могут быть взвешены по их оценочной точности с использованием обучающей выборки. Другие правила принятия решений включают байесовское усреднение, нечеткое интегрирование и теорию консенсуса (Due и др., 2012 г.).

Примеры недавних работ в литературе по дистанционному зондированию с использованием подхода множественных систем классификации включают: Briem и др. (2002 г.), Huang & Zhang (2013 г.), Zhang и др. (2015 г.), Li и др. (2012 г.), Roy и др. (2014 г.), Bigdeli и др. (2015 г.), Wu & Gao (2012 г.), Samat и др. (2014 г.), Clinton и др. (2015 г.), Han & Liu (2015 г.), Li и др. (2014 г.), Zhang & Xie (2014 г.).

3.5.2 Анализ статических данных

Статические данные относятся к данным, полученным за один период времени (динамический анализ, относящийся к анализу данных во времени, обсуждается в разделе 3.5.3). Как показано на

на рисунке выше, четыре из наиболее распространенных целей статического анализа: классификация, кластеризация, регрессия и уменьшение размерности.

Классификация применима, если общая цель состоит в том, чтобы точно распределить объекты по дискретному (обычно небольшому) набору известных классов или групп. Это распределение основано на наборе входных переменных (также называемых объясняющими переменными, факторами, переменными-предикторами, независимыми переменными, ковариатами или атрибутами). Набор данных, содержащий входные переменные и выходную переменную (соответствующие классы используются для разработки или «обучения» модели). Затем эту модель можно применить к тестовым наборам данных, которые содержат только входные переменные.

Примером может служить категоризация пикселей изображения по типам культур на основе обучающего набора данных, который содержит переменные, извлеченные из изображений (входные переменные), и наземные культуры (выходная переменная) на наборе участков. Разработанная модель классификации культур затем может быть применена к остальной части изображения, где достоверные данные недоступны.

Кластеризация применима, если целью является объединение объектов в группы или классы на основе набора входных переменных. В отличие от классификации, мы не знаем выходную переменную или классы даже в обучающем наборе данных. Поэтому нам необходимо разработать меру сходства между объектами и способ их группировки в соответствии с этим сходством. Мы можем указать количество групп (кластеров) или также сделать это неизвестным и оценить количество групп в рамках анализа. Анализ можно использовать для принятия решений об объектах, которые были сгруппированы, или для прогнозирования членства в кластере для новых объектов. Примером может служить распределение пикселей по группам на основе набора входных переменных, извлеченных из изображения. Затем эти группы можно изучить, описать и сравнить с точки зрения их характеристик.

Регрессия применима, если целью является оценка или предсказание ответной (выходной) переменной на основе набора ковариат или объясняющих (входных) переменных. Это похоже на методы классификации, но ответ является не категоричным, а непрерывным. Подобно методам классификации, модель регрессии разрабатывается или обучается на основе набора входных переменных, для которых известен ответ. Затем модель можно использовать для прогнозирования ответов для тестовых наборов данных (с входными переменными, но без переменной ответа), определения того, какие входные переменные наиболее важны для получения хороших оценок или прогнозов, или для проверки взаимосвязей между входными переменными. Примером регрессии является точная оценка или прогнозирование урожайности на основе переменных, извлеченных из изображения НЗ. В этом случае урожайность является непрерывной переменной, а не классами типов культур, описанными в приведенном выше примере классификации.

Уменьшение размерности применимо, если есть много переменных, которые можно извлечь из данных НЗ (и других источников данных), и цель состоит в том, чтобы построить небольшой набор новых переменных, содержащих всю (или большую часть) информации, находящейся в исходном (большом) наборе входных переменных. Эти новые переменные можно использовать в качестве исходных данных для других анализов или они могут быть самостоятельными конечными продуктами. Например, их можно проверять, чтобы лучше понять важные переменные, или интерпретировать как «признаки» или «индексы» (например, две переменные отражательной способности спутника объединяются для получения одной переменной индекса растительности, VI). Это неконтролируемый процесс обучения, поскольку нет переменной отклика для оценки.

3.5.3 Анализ динамических данных

Динамические данные – это данные, собранные во времени. Анализ динамических данных важен для измерения прогресса в достижении задач ЦУР. Количество периодов времени может быть маленьким или большим, в зависимости от аналитической цели. Примерами этих целей являются:

- Сравнение результатов до и после, например, сравнение качества воды или лесного покрова в результате экстремального погодного явления.
- Оценка и/или сравнение тенденций во времени, например, мониторинг ежегодных изменений в землепользовании на протяжении десятилетий или роста урожая в течение нескольких сезонов.

Если целью является сравнение результатов до и после, то типично иметь небольшое количество (НЗ и других соответствующих) наборов данных, соответствующих небольшому количеству периодов времени. Для этих типов данных есть два общих подхода.

- Проанализируйте данные за каждый период отдельно, используя методы, описанные в разделе 3.5.2, и сравните результаты.
- Возьмите разницу в значениях пикселей или объектов между данными НЗ (изображения) за два интересующих периода и проанализируйте различия, используя методы, описанные в разделе 3.5.2.
- Включите период времени в качестве ковариаты (входных данных) в методы, описанные в разделе 3.5.2.

Если цель состоит в том, чтобы оценить или сравнить тенденции с течением времени, то желательно иметь более крупные временные ряды наборов данных. Например, целью может быть мониторинг изменений в землепользовании или расширения городов в течение десятилетия, изменений в водных объектах во время и после экстремальных погодных явлений и т. д.

Существует множество подходов к анализу данных временных рядов. Fulcher и др. (2013 г.) проанализировали более 9000 алгоритмов анализа временных рядов и аннотировали коллекции из более чем 35 000 реальных временных рядов и временных рядов, сгенерированных моделями. Они разработали ресурс, который автоматизирует выбор полезных методов для классификации временных рядов и задач регрессии. Примеры типов вопросов, на которые можно ответить с помощью этого ресурса, показаны на рисунке ниже.

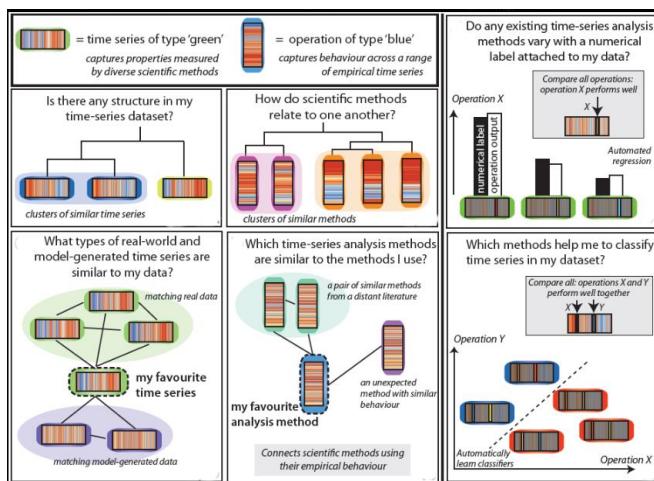


Рисунок 23: Типы вопросов анализа, на которые ответили Fulcher и др. (2013 г.)

Источник: Fulcher и др. (2013 г.). Сложный методологический взгляд на сигналы природы. <http://systems-signal.blogspot.com.au/2013/04/a-compound-methodological-eye-on.html>

Различные методы анализа динамических данных также можно классифицировать в зависимости от того, является ли целью классификация, кластеризация, регрессия или уменьшение размерности.

Классификация Для того чтобы классифицировать динамические ряды, необходимо установить, как их сравнивать. Обычный подход к сравнению кривых заключается в сопоставлении совмещения. Два метода сопоставления совмещения – это сравнение на основе примеров, которые

включают вычисление расстояния между рядами в множестве точек на протяжении ряда, и сравнение на основе признаков, которое включает сравнение набора признаков, таких как те, которые получены с помощью анализа основных компонентов.

Другие подходы к классификации кривых можно разделить на сравнительные подходы и подходы, основанные на моделях.

- Сравнительные подходы, например, кластерный анализ и анализ главных компонентов;
- Подходы на основе моделей, например, кубические сплайны, гармонический анализ и модели пространства состояний.

Эти методы и опубликованные примеры их применения к данным НЗ описаны в приложении.

Кластеризация Как описано ранее в разделе 3.5.2, общая цель состоит в объединении объектов в группы или классы на основе набора входных переменных. Ниже приведены три основные группы методов кластеризации данных временных рядов (Rani & Sikka, 2012 г.):

- Работайте непосредственно с данными временных рядов либо в частотной, либо во временной области.

Наиболее распространенные меры подобия, используемые для прямого сравнения временных рядов, включают корреляцию, расстояние между точками данных и информационные меры. Затем к этим показателям применяются методы иерархической кластеризации и k-средних.

- Непрямая работа с функциями, извлеченными из временных рядов.

Наиболее распространенные функции, извлеченные из данных временных рядов, включают точки, идентифицированные визуально, путем преобразования данных или уменьшения размерности. Наиболее распространенной мерой расстояния является Евклидово расстояние, хотя также используется расстояние Кульбака – Лейблера и геометрические расстояния.

- Работа с моделями, построенными по временным рядам.

Наиболее распространенные модели временных рядов включают модели и варианты скользящего среднего (МА), авторегрессии (AR) и модель авторегрессии скользящего среднего (ARMA), скрытые Марковские модели (HMM) и теорию нечетких множеств.

Регрессия Регрессия временных рядов направлена на прогнозирование будущего ответа на основе истории ответов от соответствующих прогнозов. Двумя распространенными методами, которые используются для этой цели, являются перечисленные выше параметрические модели временных рядов (МА, AR, ARMA и HMM), которые фиксируют временную динамику, и непараметрические сверточные нейронные сети, которые являются расширением статических нейронных сетей, адаптированных для описания данных во времени.

Другой распространенной целью регрессии временных рядов является интерполяция отсутствующих данных в пределах пространственного и временного диапазона данных. Например, облачный покров является частой причиной отсутствия данных на изображениях НЗ. Сплайны широко используются как для пространственной, так и для временной интерполяции.

Снижение размерности Общим подходом к снижению размерности временных данных является анализ основных компонентов (PCA). Этот подход описан в Приложении. Пример PCA, примененный к временному ряду значений расширенного вегетационного индекса (EVI), описан Potgieter и др. (2007 г.). EVI из 16-дневных спутниковых снимков MODIS в период посева (т. е. с апреля по ноябрь) был исследован для оценки посевных площадей пшеницы, ячменя, нута и общей

посевной площади озимых для исследуемого региона в Северо-Восточной Австралии. Метод классификации РСА и другие методы, использованные в исследовании, были отточены на основе наземных данных, собранных в исследуемом регионе. В этом исследовании анализ главных компонент использовался для сокращения временных рядов EVI в каждом пикселе в последовательности из 23 изображений в меньший набор преобразованных переменных или главных компонентов (PC), которые объясняли 90% или более временной изменчивости в рядах. В целом РСА правильно классифицировал 98% пикселей как пшеницу, ячмень, нут или области, где ничего не выращивалось. Этот пример обсуждается далее в Приложении.

3.5.4 Выбор метода

Выбор метода, используемого для анализа данных НЗ, зависит от ряда факторов:

- Характер и количество обучающих данных
- Количество калибровочных данных (достоверных наземных, в месте нахождения)
- Тип оценок и необходимых выводов
- Наличие ПО для моделирования и анализа
- Опыт команды, выполняющей моделирование и анализ

Можно сделать следующие общие выводы:

- Эмпирические методы полезны, если имеется достаточно данных для обучения и калибровки, если требуется подход, основанный на моделях, а также, если допущения моделей верны.
- Полуэмпирические методы полезны, если выполняются условия для эмпирических методов и, если есть некоторые экспертные знания о входных переменных и системе.
- Методы, основанные на физике, полезны, если есть глубокие знания о рассматриваемой системе и/или если отсутствуют калибровочные или обучающие данные.
- Модели машинного обучения полезны, если есть доступные данные для обучения моделей и оценки их соответствия, но наблюдается невозможность или нежелание делать предположения на основе моделей или разрабатывать процессы, основанные на физике.

Обзор исследований, в которых сравнивались различные статистические алгоритмы и алгоритмы машинного обучения для анализа статических данных НЗ, представлен в Приложении.

Для динамического анализа данных НЗ Австралийское бюро статистики (ABS) рекомендовало использовать модели пространства состояний для анализа динамических данных НЗ. См. Там, S-M (2015 г.). Статистическая основа для анализа больших данных для получения дополнительной информации – упоминается в Приложении.

3.6 ЭТАП 3: Оценка

Соображения, которые необходимо рассмотреть на этапе оценки, включают (1) критическую оценку результатов, (2) оценку точности результатов и (3) оценку других потенциальных предубеждений или проблем, связанных с результатами или их интерпретацией.

3.6.1 Критическая оценка результатов

Как и в случае любого статистического анализа, важно критически проанализировать результаты на предмет того, являются ли они разумными как с точки зрения статистики, так и с точки зрения предметной области. Это особенно важно для оценок, полученных на основе больших данных или нескольких источников данных. Оценки могут включать:

- Проверку общей величины результатов и выявление результатов, выходящих за пределы ожидаемых диапазонов, и т. д.;

- Проверку точности оценок, размеров интервалов прогнозирования и других показателей изменчивости результатов;
- Сравнение результатов с другими релевантными результатами;
- Оценку того, были ли соблюдены какие-либо допущения в модели или процессе моделирования.

3.6.2 Оценка точности

Точность «является относительной мерой точности оценки и учитывает систематические ошибки, также называемые предвзятостью. Таким образом, точная оценка не приводит к систематическому завышению или занижению истинного значения» (ФАО, 2016 г.).

При рассмотрении оценки точности важно признать, что усилия по выборке (точность и репрезентативность выборочных данных, используемых для обучения), а также используемые аналитические методы играют важную роль. Эти проблемы были описаны выше.

Надлежащая практика включает отчеты о подробностях этих процессов, а также сведения о соответствии модели, точности оценок или прогнозов, полученных на основе модели, и количественную оценку неопределенности этих оценок и прогнозов (Olofsson и др., 2014 г.). Сейчас эти вопросы обсуждаются более подробно.

Оценка модели

Независимо от типа модели (эмпирическая, полуэмпирическая, основанная на физике, машинное обучение) качество соответствия модели обычно оценивается путем сравнения оценочных или прогнозируемых значений, полученных из модели, с наблюдаемыми значениями, если они доступны.

Статистические модели, как правило, подвергаются дальнейшей оценке соответствия с использованием таких критериев, как AIC, BIC, коэффициентов вероятности для вложенных моделей и традиционные проверки гипотез (если таковые имеются и применимы). Там, где это возможно, оценки, полученные с помощью модели, как правило, сопоставляются с наблюдаемыми данными с использованием таких параметров, как коэффициенты ошибочных классификаций для категориальных ответов и среднеквадратическая ошибка прогноза для непрерывных ответов.

В зависимости от доступных данных типичная статистическая практика заключается в определении дополнительных наборов данных для обучения и тестирования путем случайного разделения исходных данных (например, случайного разделения набора данных на 75% обучающий набор и на 25% тестовый набор). Затем с использованием обучающих данных строится модель, которая после этого оценивается на предмет точности различными способами в тестовом наборе данных. Например, для моделей классификации коэффициенты ошибочной классификации и/или положительная прогностическая ценность (PPV) могут использоваться для оценки того, насколько хорошо обучающие данные предсказывают тестовые данные, и, следовательно, насколько хорошо обученная модель может работать на новых неизвестных выборках.

Другим широко используемым подходом являются методы перекрестной проверки. Перекрестная проверка часто используется в ситуациях, когда разделение данных, как описано выше, может значительно затруднить обучение и/или тестирование модели (например, меньшие наборы данных, где разделение приведет к тому, что размер выборки тестовых или обучающих данных будет слишком мал). Идея перекрестной проверки состоит в том, чтобы повторить процедуру проверки, описанную выше, каждый раз используя разные подмножества данных. Измеренное соответствие (например, частота ошибочной классификации) затем усредняется по этим повторам, чтобы обеспечить более точную оценку прогностических возможностей исследуемой модели. Методы перекрестной проверки, как правило, можно разделить на две группы: исчерпывающие (используются все возможные разбиения данных, например, перекрестная проверка с исключением одного или к исключений) и неисчерпывающие (не все разделы данных вычисляются – эти подходы

к перекрестной проверке можно считать аппроксимацией метода перекрестной проверки с исключением k). Выбор подхода часто просто сводится к вычислительной нагрузке.

Оценка точности данных карт

Пример методов оценки точности для анализа данных НЗ приведен в Практическом руководстве ФАО по оценке точности карт и территориальной оценке (ФАО, 2016 г., <http://www.fao.org/3/a-i5601e.pdf>). Следующая информация взята из отчета ФАО, стр. 6, 17.

При оценке точности данных карты, карта сравнивается с данными более высокого качества. Данные более высокого качества, называемые справочными данными, собираются с помощью подхода, основанного на выборке, что позволяет более тщательно интерпретировать определенные области карты. Справочные данные собираются согласованным образом и гармонизируются с данными карты для сравнения двух классификаций. Результатом сравнения являются показатели точности и скорректированные оценки территорий для каждой категории карт. Этот процесс разбит на четыре основных этапа: (i) карта, (ii) план выборки, (iii) план ответа и (iv) анализ.

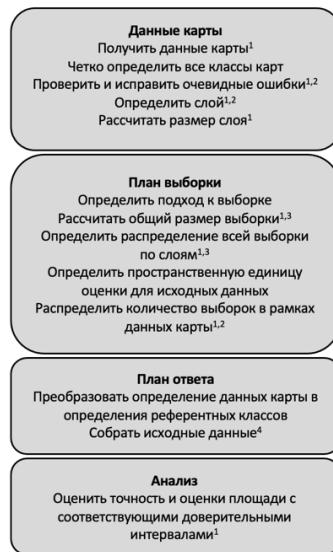


Рисунок 24: Основные четыре этапа оценки точности: получение и доработка данных карт, план выборки, план ответа и анализ. Символы показывают программное обеспечение, которое может выполнить этот этап: 1 – R, 2 – QGIS, 3 – Excel и 4 – Collect Earth.

На этапе анализа создается матрица ошибок. Матрица ошибок представляет собой перекрестную таблицу меток классов, присвоенных картой и исходными данными. Она выводится как матрица $q \times q$, где q – количество оцениваемых классов. Элементы показывают количество точек данных, которые представляют класс карты i и референтный класс j (n_{ij}). Обычно классы карт представлены в строках, а референтные классы – в столбцах. Диагональ матрицы содержит правильно классифицированные точки данных, тогда как ячейки за пределами диагонали показывают ошибки действия и бездействия.

Показатели точности рассчитываются на основе матрицы ошибок и сообщаются с соответствующими доверительными интервалами. Они включают общую точность, точность пользователя и точность производителя. Общая точность представляет собой долю правильно классифицированной территории и, таким образом, относится к вероятности правильной классификации случайно выбранного места на карте (см. уравнение 3). Точность пользователя – это доля территории, классифицируемой как класс i , который также является классом i в справочных данных (см. уравнение 4). Этот показатель сообщает пользователям вероятность того, что определенная область карты класса i также является классом на земле. Точность производителя –

это доля территории, которая относится к референтному классу j и также относится к классу j на карте (см. уравнение 5). Это вероятность того, что класс j на земле отображается как тот же класс.

$$A = \sum_{j=1}^q p_{jj} \quad (3)$$

$$U_i = p_{ii}/p_{i\cdot} \quad (4)$$

$$P_j = p_{jj}/p_{\cdot j} \quad (5)$$

Для всех трех показателей точности также необходимо рассчитать доверительные интервалы. Формула дисперсии представлена в уравнениях 5, 6 и 7 в работе Olofsson и др. (2014 г.), а 95% доверительный интервал можно рассчитать, умножив квадратный корень из дисперсии на 1,96.

ФАО применила эти этапы оценки точности для измерения площади лесов и изменения площади лесов. Эти меры важны для стран, предъявляющих требования к отчетности, чтобы получить доступ к платежам, основанным на результатах, за сокращение выбросов в результате обезлесения и деградации лесов.

Интерпретация оценок точности

Для оценки точности карты достоверность данных зависит от предназначения карты и должна оцениваться на индивидуальной основе. Следующее обсуждение взято из отчета ФАО (2016 г.), стр. 20, 21.

Например, в случае изменения земного покрова необходимо учитывать точность изменений, а не точность двух отдельных карт земного покрова. Даже в том случае, если обе карты земного покрова содержат высокоточные показатели за один момент времени, они не содержат никакой информации о точности для классов изменений. Необходимо провести новый анализ изменений с использованием изображений, полученных с помощью дистанционного зондирования, а не сравнивать карты, составленные в разное время.

Общая точность карты не всегда отражает точность отдельных классов (GFOI, 2013 г.). Высокая общая точность карты не гарантирует высокой точности в отношении потери лесов. Поэтому необходимо учитывать как точность производителя, так и точность пользователей по всем отдельным классам.

Кроме того, общий размер выборки, число страт и распределение общего размера выборки по стратам могут благоприятствовать одному показателю точности по сравнению с другим. Точность обычно выше для стабильных классов, чем для классов изменений. Кроме того, точность варьируется в зависимости от ландшафта. При оценке значения точности необходимо учитывать этапы оценки точности, описанные в практическом руководстве ФАО.

Отчет о результатах оценки точности

Снова выдержка из отчета ФАО (2016 г.), стр. 23:

При представлении результатов оценки точности отчет должен включать не только оценки точности, скорректированную площадь и соответствующие доверительные интервалы, но и предположения, вытекающие из нескольких элементов оценки точности. Допущения могут влиять на уровень точности и включают, но не ограничиваются следующим:

1. минимальной единицей картирования и единицей пространственной оценки
2. планом выборки
3. определением леса
4. источником справочных данных

5. доверительным уровнем, используемым для расчета доверительных интервалов (обычно 95%).

Оценки всегда должны сообщаться с соответствующими доверительными интервалами. Кроме того, рекомендуется представлять матрицу ошибок в виде оценочных долей территории вместо абсолютной выборки.

Для получения подробной информации о практическом применении оценки точности см. пример глобального изменения лесов (GFC) в Практическом руководстве ФАО по оценке точности карт и территориальной оценке 2016 года.

3.6.3 Другие проблемы

Другие опасения по поводу результатов анализа данных НЗ могут включать:

- систематическую ошибку в данных, используемых для получения оценок модели
- предположения, лежащие в основе принятой модели
- влияние выбора метода оценки интересующих величин
- возможную неверную интерпретацию результатов
- погрешность оценок площади, вычисленных непосредственно по классифицированным изображениям

Первые два пункта были рассмотрены ранее в этой главе, а последний пункт обсуждался выше. Обеспокоенность по поводу выбора метода оценки важна, поскольку она поднимает вопрос о допустимой субъективности результатов. Учитывая широкий спектр классификаторов изображений и аналитических методов, существует опасение, что их можно «настроить» для получения «правильных» результатов, например желаемой пропорции между классами, и, следовательно, чтобы аналитик мог выбрать/настроить алгоритм для получения значения, которое лучше соответствует априорному убеждению.

Еще одна проблема заключается в том, что точность коррекции смещения зависит от размера выборки используемых полевых данных. Следствием этого является то, что добавленная стоимость дистанционного зондирования обычно пропорциональна усилиям, затраченным на полевые исследования, за исключением некоторых классов, для которых идентификация на изображении особенно надежна.

Эти проблемы в значительной степени могут быть решены путем следования руководствам по передовой практике – см., например, руководство GEOSS, составленное Gallego и др. (2008 г.) – кроме того, путем продвижения прозрачной отчетности, открытых данных, открытых моделей и результатов, а также благодаря тиражированию результатов независимыми группами.

3.7 Резюме

В этой главе описан ряд методологий для анализа данных НЗ. После обсуждения типа оценок, которые могут быть получены из данных НЗ, основное внимание было уделено трем основным этапам – предварительной обработке данных, методам анализа данных и критической оценке результатов анализа.

Был предоставлен ряд ссылок, указывающих на актуальность описанных методов и их практическое применение к данным НЗ. Области применения варьируются от сельскохозяйственной и экологической статистики до других областей, таких как определение качества воды и рост городов. Актуальность этих оценок для целей в области устойчивого развития на период до 2030 года находится в центре внимания главы, как описано в разделе 3.2.

Также были предоставлены ссылки на технические детали методов, описанных здесь. Более подробная информация о выбранных методах представлена в Приложении.

В дополнение к ссылкам в главе также выделены тематические исследования для каждого из трех основных шагов. К ним относятся детали практического подхода к предварительной обработке данных (разработанный и используемый DSITI (Австралия) уже более десяти лет), модель пространства состояний для анализа временных рядов изображений дистанционного зондирования (рекомендована Статистическим бюро Австралии) и руководящие принципы передовой практики для оценки точности на основе рекомендаций ФАО (2016 г.) и GEOSS (2008 г.). Более подробные тематические исследования, в которых используются методы, описанные в этой главе, представлены в главе 4.

4. Пилотные проекты и тематические исследования

Под эгидой Целевой группы был разработан ряд экспериментальных проектов, охватывающих целое множество областей применения. Эти проекты согласуются с задачами и стратегиями Целевой группы; способствуют совершенствованию официальной статистики за счет использования спутниковых изображений, а также поддерживают мониторинг ЦУР.

Пилотные проекты можно рассматривать как примеры того, как Целевая группа работает над своими четырьмя конкретными задачами:

- Определить надежные и точные статистические методы для оценки интересующих величин;
- Предложить подходы к сбору репрезентативных обучающих и проверочных данных надлежащего качества;
- Исследовать, разработать и внедрить методы оценки предлагаемых моделей, включая меры точности и согласия;
- Разработать стратегии повторного использования и адаптации алгоритмов по различным темам и содействовать внедрению больших объемов данных.

Естественно, прогресс, достигнутый на сегодняшний день, не является однородным. Пилотные проекты показывают разные уровни развития. В этой главе представлены три наиболее зрелых проекта:

Пилотный проект 1: Применение данных спутниковых изображений для производства сельскохозяйственной статистики, осуществляется Австралийским бюро статистики (ABS).

Пилотный проект 2: Исследование по оценке товарных запасов Skybox, возглавляемое СОООН и проводимое Google.

Пилотное исследование 3: Исследование по использованию спутниковых изображений для расчета статистических данных о земном покрове и землепользовании, проведенное DANE, Колумбия.

В каждом отчете описываются методы обработки, источники данных, критерии отбора и вопросы качества, связанные с данными, а также анализы и результаты, полученные на момент составления этого отчета.

В дополнение к пилотным проектам в эту главу включены пять тематических исследований, дополняющих пилотные проекты Целевой группы. Тематические исследования заключаются в следующем:

Тематическое исследование 1: оценка внедрения различных методов машинного обучения для анализа данных НЗ с целью классификации сельскохозяйственных культур, проведенная Департаментом науки, информационных технологий и инноваций (DSITI) Австралии и Квинслендским технологическим университетом (QUT) Австралии.

Тематические исследования 2 и 3: две опубликованные интеграции анализа данных НЗ по оценке урожайности.

Тематические исследования 4 и 5: разработаны два полностью действующих метода анализа данных НЗ для получения официальной статистики. Они были разработаны, соответственно, Национальным статистическим управлением Канады и DSITI Австралии.

4.1 ЭКСПЕРИМЕНТАЛЬНЫЙ ПРОЕКТ 1: Применение данных спутниковых изображений для производства сельскохозяйственной статистики (Австралия)

Экспериментальное исследование Австралийского бюро статистики (ABS) по применению данных спутниковых изображений в производстве сельскохозяйственной статистики приводит следующие результаты исследования: «(...) оценка применимости статистической методологии для классификации спутниковых данных об отражательной способности по типу и оценке продукции растениеводства». Этот проект был основан на эмпирическом подходе. Полный текст документа можно получить по запросу у Австралийского бюро статистики (Marley и др. 2014 г.).

Характеристики данных

ABS получило данные изображений Landsat 7 от Geoscience Australia (GA), поскольку это был активный спутник Landsat, собирающий данные в течение периода, за который у них были точные наземные данные. Поскольку пилотный проект был проведен, в качестве источников рекомендуется использовать данные Landsat 8 или Sentinel 2. Данные изображений Landsat в Австралийском кубе геонаучных данных (AGDC) хранятся в наборах тайлов широты и долготы с шагом в один градус. Каждый тайл имеет размер примерно 100 км x 100 км, при этом каждый пиксель представляет собой участок земли размером 25 x 25 м. С каждым пикселием связано шесть измерений коэффициента отражения, соответствующих шести различным диапазонам частот электромагнитного спектра, а именно красному, зеленому, синему, ближнему инфракрасному (NIR), коротковолновому инфракрасному (SWIR) 1 и SWIR 2. Landsat 7 повторно посещает один и тот же участок земли в среднем каждые 16 дней, а это означает, что большинство пикселей имеют набор из шести измерений отражательной способности каждые 16 дней.

ABS получило данные спутниковых снимков от AGDC за 2010–2012 календарные годы. Данные Landsat использовались для классификации сельскохозяйственных культур. Объем отсутствующих данных на изображениях Landsat 7 является ключевой проблемой качества, которая описана в разделе 4.1.4. Для будущих проектов эту проблему можно решить, используя данные Landsat 8 или Sentinel 2. Поскольку имеется значительный объем недостающих данных, возможное направление будущих исследований заключается в изучении возможных методов корректировки или условного исчисления недостающих данных. Возможным решением является использование подходов асимиляции модельных данных, описанных в Главе 2.

Несколько продуктов данных MODIS доступны в Государственном объединении научных и прикладных исследований (CSIRO), и те из них, которые представляли особый интерес для пилотного исследования, перечислены в таблице 6. Данные из продукта MOD13Q1.005 использовались для изучения возможности обнаружения урожая, в ходе которого измерялось наличие или отсутствие культур независимо от их типа. Продукт MOD13Q1.005 содержит слой расширенного вегетационного индекса (EVI) и два основных интересующих слоя качества; а именно, уровень надежности пикселей и уровень качества индекса растительности (VI).

EVI является примером полуэмпирического типа алгоритма, упомянутого в главах 2 и 3, поскольку известные спектральные каналы имеют отношение к хлорофиллу и, таким образом, измерение фотосинтеза у растений, а также ближайший пик инфракрасного отражения, вызванный здоровыми листьями.

Продукт	Описание продукта
МОД09А1.005	500 млн пикселей 8-дневные композиты ²⁰ Каналы от 1 до 7 плюс слои качества
МОД09ГА.005	500 млн пикселей Ежедневно

²⁰ Продукты, описываемые как 8 или 16-дневные композиты, показывают, что данные, предоставляемые для каждого пикселя изображения являются лучшим качеством измерения за 8 или 16 дней. В этих случаях включается дополнительная переменная, указывающая точный день года, в который производилось измерение данных.

Каналы от 1 до 7 плюс слои качества	
MOD09Q1.005	250 м пикселей 8-дневные композиты Канал 1 и канал 2 (красный и NIR)
МОД11А1.005	1000 млн пикселей Ежедневно Температура/коэффициент излучения поверхности земли
МОД11А2.005	1000 млн пикселей 8-дневные композиты Температура/коэффициент излучения поверхности земли
MOD13Q1.005	250 млн пикселей 16-дневные композиты Индексы растительности (EVI, NDVI) плюс слои качества

Таблица 6: Продукты данных MODIS, использованные в пилотном исследовании ABS.

Обработка данных

В этом пилотном исследовании разрабатываются различные методы обработки для подготовки данных с целью использования в точной классификации спутниковых изображений по различным классам сельскохозяйственных культур и получения качественных статистических данных о посевных площадях. Технические задачи заключались в подготовке и исследовании обучающих данных для классификации типов культур, а также в подготовке обучающих данных для обнаружения этих культур.

Обучающие данные – это данные, используемые для построения классификационных моделей; они являются результатом интеграции данных спутниковых снимков и данных наземной разведки, а также сопоставления результатов измерений поверхностного отражения с помощью спутников с классами земного покрова или землепользования.

Начиная с набора достоверных наземных данных из технических задач Департамента науки, информационных технологий и инноваций Квинсленда (DSITI), подготовка обучающих данных для классификации типов сельскохозяйственных культур состояла из двух частей. Во-первых, с определения границы поля и, во-вторых, с оценки сроков посева и сбора урожая на уровне поля.

Границы поля были определены вокруг наблюдаемых точек широты и долготы, чтобы зафиксировать изменения поля. Ручная проверка данных спутниковых изображений и нанесение границ с использованием геопространственного программного обеспечения были наиболее точным и эффективным по времени способом определения границ поля для этого мелкомасштабного технико-экономического обоснования.

Для каждого пикселя в полученных вручную границах поля, были извлечены измерения отражательной способности Landsat 7 для всех временных точек, доступных с 2010 по 2012 год, с использованием R-кода, разработанного внутри компании. На основе этого интегрированного набора данных следующей задачей было определить даты посева и сбора урожая, которые определяют вегетационный период для каждого поля.

Оценка сроков посева и сбора урожая на уровне поля была следующей;

- Расчёт среднего значения EVI на уровне пикселей для каждого поля в каждый наблюдаемый момент времени;
- Подбор сглаживающего сплайна к нерегулярным времененным рядам медианных значений EVI на уровне поля; затем даты, связанные с ближайшим локальным минимумом по обе

стороны от даты, когда были обнаружены данные о поверхности земли, были помечены как даты посева (слева) и сбора урожая (справа).

С целью обеспечить точную оценку сроков посева и сбора урожая, был включен ряд правил для полей, у которых достоверным наземным данным соответствовало недостаточно данных спутниковых изображений (из-за отсутствия данных, низкого качества изображений и облачного покрова). Из 962 проверенных мест этот автоматизированный метод определил сроки посева и сбора урожая на 414 площадях. Эти 414 полей включали только поля с сезонным характером отражения, показывающим отчетливые пики и спады. Для будущих исследований можно рассчитать дисперсию подобранного сплайна.

Чтобы создать окончательный набор обучающих данных, эти предполагаемые даты посева и сбора урожая были использованы для разделения промежуточного интегрированного набора данных только на те измерения отражательной способности пикселей, которые соответствовали вегетационному периоду наблюданной культуры. Также был создан второй обучающий набор данных, который содержал сглаженные медианные значения EVI через равные промежутки времени в течение предполагаемого вегетационного периода каждого наблюданного поля.

Был проведен исследовательский анализ обучающих данных, чтобы определить подходящие ковариаты, которые могли бы повысить прогностическую способность модели классификации сельскохозяйственных культур, с использованием повторного дисперсионного анализа EVI и других измерений. Повторный дисперсионный анализ на уровне пикселей со случайными эффектами для класса культуры и поля был проведен для проверки различий измерений отражательной способности и EVI между классами культур (а именно, кукурузой, сорго, ячменем и пшеницей). При отсутствии эффектов случайного поля дисперсионный анализ показал, что средние измерения коэффициента отражения значительно различались между типами культур. Однако при включении эффектов случайного поля эти различия ис��али. В то время как внутри полей и между ними наблюдались значительные различия, различия между культурами не были статистически значимыми.

Прочность взаимосвязей между характеристиками почвы и ландшафта, доступными в данных CSIRO Soil and Landscape Grid (SLG), и значениях EVI были исследованы с помощью различных методов визуализации, включая графики плотности, прямоугольные графики и графики рассеяния. Визуализации показали, что недостаточно доказательств, чтобы указать, что переменные SLG имеют существенную связь с измерениями спутниковой отражательной способности. Случайные леса также рассматривались для выбора переменных. В результате этого краткого исследования было обнаружено лишь небольшое количество значимых взаимосвязей. Например, для посевов пшеницы требуются более глубокие почвы с более высоким содержанием глины. Однако результаты показали, что включение переменных почвы и ландшафта на самом деле увеличило ошибку прогноза; следовательно, это не рассматривалось далее.

Ряд фенологических характеристик был получен из сглаженных кривых EVI на уровне поля для использования в моделях классификации сельскохозяйственных культур, включая продолжительность цикла роста сельскохозяйственных культур в днях, количество дней, затраченных на достижение максимального EVI с момента посева, области под кривой в различных фазах роста и в пределах полевых показателей распределения EVI, таких, как диапазон, максимум, минимум, экспесс, асимметрия и дисперсия. Основным ограничением характеристик кривой роста сельскохозяйственных культур для статистического моделирования является потенциально высокая степень мультиколлинеарности между переменными. Мультиколлинеарность можно преодолеть с помощью методов уменьшения размерности, таких как анализ основных компонентов (PCA). Также была опробована альтернативная методология, функциональный анализ данных, в результате которого из всего (сглаженного) ряда EVI были созданы функциональные формы для вегетационного периода. Этот подход позволил избежать проблем выбора переменных модели подходящих фенологических характеристик и проблемы мультиколлинеарности.

Следующие обучающие данные были подготовлены для обнаружения урожая. Все началось с построения модели обнаружения урожая, которая служила основой для создания маски урожая. Для создания маски урожая требуются достоверные наземные данные, содержащие участки, которые использовались и не использовались для урожая. Северо-западный регион штата Виктория был выбран в качестве пилотного региона для этого исследования, для которого достоверные наземные данные были доступны из Базы данных основных участков почвенного покрова Австралии (AGCRDS) и сельскохозяйственной переписи ABS 2010-11.

Набор наземных данных включал 18 уникальных участков в AGCRSD в северо-западном регионе Виктории и девять участков от респондентов сельскохозяйственной переписи 2010–2011 годов в изучаемой области. Дополнительные участки, которые использовались для несельскохозяйственного земледелия (например, водоемы, национальные парки, леса и т. д.), требовались для получения сбалансированных наземных данных, и они были выбраны вручную на глаз с помощью Google-карт. Границы данные о широте и долготе выбранных участков были записаны в набор достоверных наземных данных вместе со статусом урожая и временем года; а также год, когда были собраны наземные данные, в котором состоялось наземное наблюдение.

Выбор данных

Для достижения целей проекта были включены различные источники данных; спутниковые снимки (в основном MODIS и Landsat); достоверные наземные данные (сельскохозяйственная перепись, основные участки и данные придорожных наблюдений, собранные DSITI). В будущей работе эти наборы данных могут быть дополнены дополнительными источниками данных, такими как Sentinel-2, VIIRS, Sentinel-3, данные радара с синтезированной апертурой от Sentinel-1, данные SLG, а также метеорологические и климатические данные.

Тем не менее, существует ключевой вопрос, касающийся наличия и надлежащего качества источников данных НЗ; свободно и коммерчески доступные данные имеют ограничения, связанные с пропущенными значениями, периодичностью и качеством. Были рассмотрены два пути преодоления этих проблем. Во-первых, это проверка и стандартизация. Это включало сравнение данных из одного и того же источника в разное время и/или сравнение между несколькими источниками. Данные AGDC Landsat для Австралии были предварительно обработаны и откалиброваны в соответствии с одним и тем же стандартом на протяжении всего времени. Данные MODIS CSIRO были аналогичным образом предварительно обработаны, однако другие источники спутниковых изображений должны быть стандартизированы перед использованием в области анализа обнаружения и классификации культур. Второй проблемой был потенциал. Даже свободно доступные данные и инструменты требуют квалифицированного персонала (например, специалистов по данным) для надлежащего извлечения огромного количества содержащейся в них информации. Тем не менее, этот процесс может быть весьма ресурсоемким, и его необходимо планировать при любой работе с большими данными.

4.1.4 Вопросы качества

Отсутствующие данные были ключевой проблемой качества для этого пилотного проекта. Существует два основных источника недостающих данных:

1. Изображения Landsat 7 изобилуют полосами отсутствующих данных, вызванными неисправностью корректора линии сканирования на борту спутника с мая 2003 года. Полосы равномерно расположены по краю сканирования Landsat и не должны коррелировать с типом сельского хозяйства, которое ведется в области пораженных пикселей; следовательно, это не должно приводить к статистической погрешности, за исключением случаев уменьшения эффективного размера выборки. Это не проблема для других спутников, в частности для Landsat 8, который в настоящее время используется и получает данные, и для Sentinel-2, который в настоящее время собирает более частые данные с разрешением 10 м.

2. Geoscience Australia (GA) предоставила качественный слой с данными спутниковой съемки, рассматривая любые пиксели, помеченные как некачественные и как непригодные. Основными причинами плохого качества являются облачный покров или тень от облаков. Этот подход предотвращает низкое качество пикселей, влияющих на результаты.

Также существуют значительные различия в схемах отражения со спутников между различными полями, выращивающими один и тот же вид сельскохозяйственных культур. Получение дополнительных переменных-предикторов, которые объясняют эти вариации от одной области к другой, может помочь повысить точность прогнозов, получаемых с помощью классификационной модели, однако это оказалось серьезным препятствием для получения более точных статистических результатов. Эта проблема связана с эмпирическим подходом, однако потенциально ее можно было бы решить путем использования биофизических подходов, которые не входили в рамки данного пилотного проекта.

4.2 ПИЛОТНЫЙ ПРОЕКТ 2: Оценка сырьевых запасов Skybox – теперь часть Planet (UNSD и Google)

В пилотном проекте Google изучалась оценка запасов сырой нефти с использованием аналитики, полученной на основе спутниковых изображений, в качестве средства повышения экономической прозрачности.

Пилотный проект был специально направлен на количественную оценку объемов базовых товаров, а именно сырой нефти, с использованием спутниковых изображений. Торговля этим товаром, хотя и имеет жизненно важное значение для глобальной системы и занимает центральное место в экономике многих развивающихся стран, часто недоступна или неточна.

Измеряя тени на резервуарах с сырой нефтью с плавающей крышкой, Skybox попытался оценить объем запасов, присутствующих в данном резервуаре. Это было сделано с помощью довольно простой тригонометрии, принимая во внимание широту и долготу цели, время суток и угол съемки спутника.

Чтобы оценить точность основной методологии, Google провел анализ нефтехранилища в Кушинге, штат Оклахома, США. Это крупнейшее нефтехранилище в мире, по которому EIA публикует еженедельные данные о хранящихся там запасах нефти. Анализируя разницу между оценками Google и официальной отчетностью, можно оценить точность подхода.

Характеристики данных

Панхроматический	450 – 900 нм
Синий	450 – 515 нм
Зеленый	515 – 595 нм
Красный	605 – 695 нм
NIR	740 – 900 нм

Панхроматический GSD	90 см в надире
Мультиспектральный GSD	2 м в надире
Ширина полосы съемки	8 км в надире
Формат файла	16-битный файл Geo TIFF

Таблица 7: Характеристики изображения

Цвет	Панхроматический
Продолжительность	До 90 секунд
Частота кадров	30 кадров в секунду
GSD	1,1 м в надире
FOV	2 км на 1,1 км
Формат файла	MPEG-4 (кодировка H 264)

Таблица 8: Характеристики видео

Обработка данных

Обычные проблемы, такие как облачный покров, были представлены и решены путем сочетания ручной и автоматизированной обработки. Дальнейшие подробности о методах обработки данных не были включены из-за коммерческой конфиденциальности, связанной с проектом.

Выбор данных

Данных было достаточно, чтобы соответствовать цели проекта. В данном случае использовался один источник; Снимки SkyBox (теперь Planet) с субметровой точностью.

Данные, используемые Google, сосредоточены на мониторинге конкретных целей, имеющих большое экономическое значение, в частности, на крупных нефтеперерабатывающих заводах, крупных городских районах, шахтах и морских портах.

Проблемы с качеством

Google использовал изображения Skybox. Что касается предварительной оценки запасов, Skybox использовала запатентованную комбинацию полностью автоматизированных и ручных процессов. Ошибка в измерениях исследуется на постоянной основе, но потенциальные источники включают относительно высокий угол наклона солнца в летние месяцы, методологию обработки изображений, отношение сигнал/шум, особенности оборудования и инфраструктуры, а также угол съемки спутников.

В некоторых случаях в ходе сбора данных были пропуски из-за облачности. Однако с полной группировкой из 24 спутников Skybox сможет наблюдать за любым местом на Земле от пяти до семи раз в день. Это может значительно помочь решить проблемы с пропусками в данных.

4.3 ПИЛОТНЫЙ ПРОЕКТ 3: Использование спутниковых изображений для расчета статистических данных о земном покрове и землепользовании (Колумбия)

Национальный административный департамент статистики Колумбии (DANE) провел пилотный проект с целью предложить метод расчета показателя ЦУР «Соотношение темпов землепользования и темпов прироста населения» и пытается получить оценку эффективности землепользования путем мониторинга взаимосвязи между землепользованием и приростом населения. (SDSN, 2016 г.):

«(...) этот показатель определяет и отслеживает взаимосвязь между землепользованием и ростом населения. Он информирует и позволяет лицам, принимающим решения, отслеживать рост городов и управлять им в различных масштабах, а также расширяет их возможности по повышению эффективности землепользования. В целом, он гарантирует, что ЦУР должным образом охватывают более широкие аспекты пространства и земли, а также обеспечивают основу для реализации нескольких других целей, в частности, в области здравоохранения, продовольственной безопасности, энергетики и изменения климата».

Цели проекта:

- Оценить возможность применения методов дистанционного зондирования для оценки уровня землепользования и пространственного анализа на основе географических информационных систем (GIS) с целью расчета взаимосвязи между землепользованием и ростом населения в городских районах;
- Оценить методологию путем ее применения к столичному району Барранкильи с помощью спутниковых изображений и данных о населении за 2005, 2010 и 2015 годы.

Спецификации данных

Для расчета показателя был выбран временной интервал в пять лет. Этот критерий считался подходящим, чтобы увидеть эволюцию населения и иметь адекватный охват изображениями. В анализ включены 2005, 2010 и 2015 годы.

Критерии выбора изображений:

- Доступность
- Облачность менее 10% на изучаемой территории
- Один и тот же месяц для каждого года, чтобы свести к минимуму изменения, вызванные сезонами дождей или засухи, фазами урожая и прочим
- Полное покрытие мегаполиса, чтобы избежать мозаики изображений
- Похожие спектральные каналы.

Использовались бесплатные изображения Landsat, которые были загружены с веб-сайта Earth Explorer Геологической службы США (USGS) (<http://earthexplorer.usgs.gov/>). Район исследований Landsat имеет следующие значения номера колонки и номера строк (path-row) 9-52. В таблице 10 показаны миссия, дата и имя файла для каждого изображения, использованного в пилотном проекте.

Path-Row	Миссия	Дата съемки	Файл
9-52	Landsat 7	7 января 2005 г.	LE70090522005007ASN00
9-52	Landsat 7	8 февраля 2005 г.	LE70090522005039EDC00
9-52	Landsat 5	29 января 2010 г.	LT50090522010029CHM01
9-52	Landsat 8	11 января 2015 г.	LC80090522015027LGN00

Таблица 10: Снимки Landsat, включенные в пилотный проект

На 2005 год было собрано две сцены. Это было продиктовано необходимостью улучшения изображения. Вся процедура описывается в разделе, посвященному обработке.

Обработка данных

Методы обработки включали два этапа; предварительная обработка изображений и обнаружение изменений (сегментация и классификация изображений). Предварительная обработка изображений включала процедуры подготовки изображений к процессу классификации; заполнение пробелов (для изображения 2005 г.), масштабирование радиометрического разрешения (для изображения 2015 г.), картографическая проекция, создание подмножества изображений и наложение слоев. Чтобы заполнить пробелы, был использован алгоритм Gap-Fill USGS для улучшения изображений 2005 года Landsat 7 ETM+ ввиду отсутствия корректора линий сканирования (SLC). Масштабирование радиометрического разрешения было выполнено для снимка Landsat 8 2015 года.

Были созданы подмножества изображений, охватывающих буфер радиусом в 500 метров вокруг столичного района Барранкилья. Это расстояние является компромиссом между охватом окружающих территорий и пространственным ограничением изображения Landsat 5 2010 года, которое меньше, чем у других миссий. Переход от Landsat 5 к Landsat 7, а затем к Landsat 8 включал добавление новых каналов (наложение слоев), а некоторые ранее определенные были разделены. Тем не менее, большинство из них сопоставимы.

Вторым выполненным процессом было обнаружение изменений. Оно использовалось для постклассификационного сравнения. В этом методе каждая дата исправленных изображений классифицируется независимо, чтобы соответствовать общей схеме типов земель (равное количество и тип классов земельного покрова). Затем полученные карты земного покрова накладываются друг на друга и сравниваются попиксельно. В результате получается карта изменения земного покрова. Это попиксельное сравнение также можно обобщить в матрице изменений «от-до», также называемой матрицей перехода. Матрица изменений «от-до» показывает все возможные изменения земного покрова в соответствии с исходной схемой классификации.

Классификация изображений – это процесс, используемый для создания тематических карт на основе изображений. Поскольку одной из целей пилотного проекта было выявление городских поселений и учитывая, что их спектральный отклик очень похож на отклик местности, лишенной растительности или местности с минимальной растительностью, поэтому был использован метод OBIA для включения дополнительных признаков к спектральной функции и улучшения тематической точности. Для выполнения процесса использовался INTERIMAGE с открытым исходным кодом (без оплаты лицензии). Изображение, которое нужно классифицировать, и семантическая сеть (иерархические объекты, которые будут идентифицированы на изображении) были введены в INTERIMAGE в качестве входных данных. Затем были определены методы и операторы, позволяющие генерировать объекты и их классифицировать.

Рисунок 25 описывает компоненты процесса интерпретации на INTERIMAGE. Система реализует определенную стратегию управления интерпретацией, руководствуясь структурированной моделью знаний через семантическую сеть (иерархические сущности, которые будут идентифицированы на изображении). Управление интерпретацией выполняется ядром системы, которое использует в качестве входных данных набор изображений с географической привязкой, слои SIG, цифровые данные высот или другие данные с географической привязкой. Через интерпретацию сцены входные данные обрабатываются с помощью операторов «сверху вниз» и «снизу вверх».

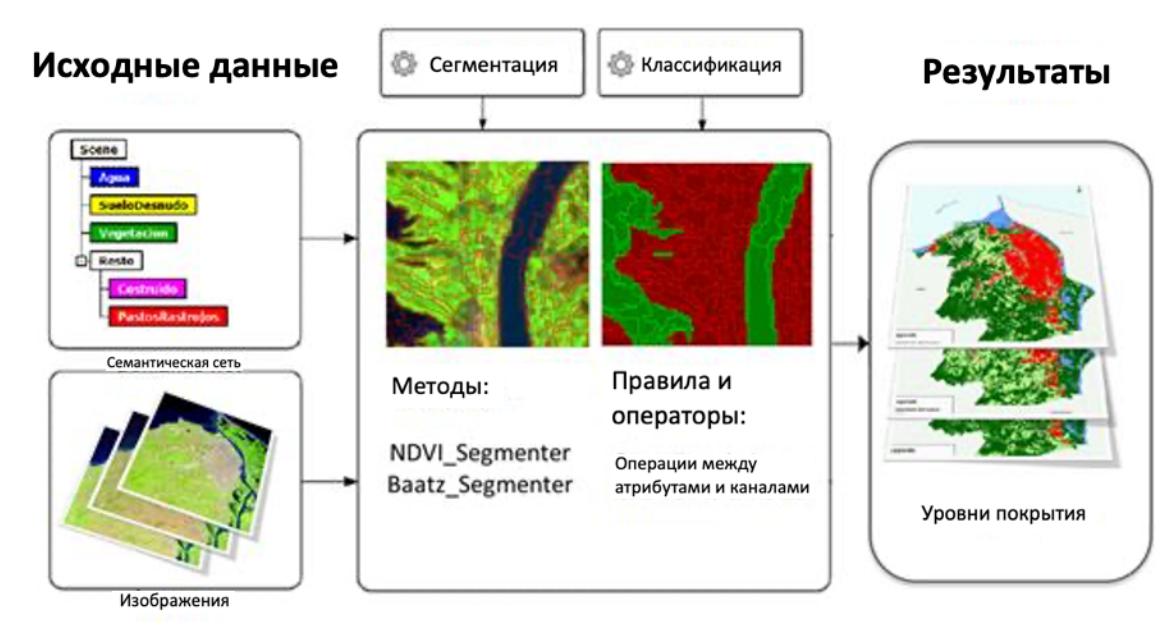


Рисунок 25: Блок-схема INTERIMAGE для объектной классификации.

Операторы сверху вниз отвечают за разбиение сцены на области, рассматриваемые как объекты (сегментация). Этот процесс создает объекты путем группировки похожих пикселей. Он начинается с того, что каждый пиксель формирует один объект изображения, который последовательно объединяется, чтобы сформировать более крупные объекты. Числовой индикатор Нормализованный разностный вегетационный индекс (NDVI) и алгоритмы TA Baatz Segmenter использовались для получения сегментации изображения.

После определения алгоритмов сегментации были введены решающие правила для каждого узла семантической сети. Эти правила позволяют отфильтровывать из набора объектов те, которые принадлежат интересующему классу. В решающем правиле можно вычислить множество атрибутов на основе спектральных значений, формы, текстуры и топологических характеристик сегментов изображения. Эти атрибуты можно использовать для выбора объектов в наборе с заданным пользователем порогом. Объекты, которые не находятся в пределах порога, оцениваются в другом классе в соответствии с порядком, определенным в семантической сети. Если объект соответствует критериям более чем одного класса, он относится к более надежному классу.

Операторы сверху вниз отвечают за разбиение сцены на области, рассматриваемые как объектные гипотезы (сегментация). Это предварительная классификация, которая определяет сегменты, потенциально принадлежащие к каждому классу. Операторы «снизу вверх» уточняют классификации, полученные на этапе «сверху вниз» (подтверждая или отвергая) и разрешая возможные пространственные конфликты между ними. В конце процесса интерпретации гипотезы становятся проверенными экземплярами объекта (классификация).

Изображение, которое нужно классифицировать, и семантическая сеть (иерархические объекты, которые будут идентифицированы на изображении) вводятся в INTERIMAGE (промежуточное изображение) в качестве входных данных. Затем определяются методы и операторы, позволяющие генерировать объекты и их классификацию.

Семантическая сеть проекта была создана на основе предыдущих знаний об изучаемой территории и с учетом классификации земного покрова на базе снимков Landsat из проекта GeoCover²¹.

²¹ <http://www.mdafederal.com/geocover>

Наконец, числовой индикатор Нормализованного разностного вегетационного индекса (NDVI) и алгоритмы TA Baatz Segmenter использовались для получения сегментации изображения.

Выбор данных

Несмотря на некоторые проблемы с качеством, данные Landsat подходили для целей проекта, потому что они были в свободном доступе и существует несколько инструментов для решения проблем с качеством. Для будущих проектов с использованием источников данных НЗ, таких как Landsat 8 и Sentinel-2, можно решить проблемы с качеством, связанные с данными Landsat 7.

Вторым источником данных были данные о населении. Учитывая, что 2005 г. был последним годом проведения переписи населения в Колумбии, использовались прогнозы численности населения. Данный факт может повлиять на качество результатов, но это наиболее распространенный и доступный способ получения информации о населении в периоды между переписями. Этот источник данных был достаточным для достижения целей проекта.

Критерии отбора данных были связаны со свободным доступом и имеющимися возможностями. В этом смысле они были уместны. Возможный поиск дополнительных источников может быть выполнен в ходе будущей работы с использованием, например, изображений MODIS, Sentinel-2, Landsat 8, Sentinel-3 VIIRS и новых инструментов обработки, таких как Google Earth Engine.

Проблемы с качеством

После использования алгоритма Gap-Fill USGS улучшенные изображения Landsat 7 показали достаточное качество для целей исследования. Матрица ошибок была рассчитана с использованием изображений Google Earth и ортофотоснимков для информации. Эта матрица, как упоминалось ранее в главе 3, позволяет оценить качество классификации.

4.4 ПРИМЕР 1: Применение методов машинного обучения к данным спутниковых изображений при производстве сельскохозяйственной статистики (Австралия)

В качестве расширения пилотного проекта ABS для целевой группы, исследователи из Квинслендского технологического университета (QUT) в Австралии применили ряд методов машинного обучения по отношению к проблеме классификации сельскохозяйственных культур, используя те же данные, которые были получены для пилотного проекта №1. Цель состояла в том, чтобы оценить простоту применения методов и сравнить полученные оценки.

Методы машинного обучения включали нейронные сети, бустинг деревьев регрессии и регуляризованную полиномиальную регрессию. Эти методы описаны в Приложении. В качестве форм выбора переменных использовались два типа регуляризации: регуляризация Риджа существенно уменьшала коэффициенты неэффективных переменных, а регуляризация Лассо уменьшала их до нуля и, следовательно, эффективно исключала их из анализа.

Характеристики данных

В этом тематическом исследовании использовались данные Landsat из пилотного проекта ABS. Спецификации данных описаны в разделе 4.1.1.

Переменной ответа был тип культуры. Набор достоверных наземных данных включал семнадцать (17) типов культур: голая почва, ячмень, хлопок, леукена, люцерна, кукуруза, просо, маш, овес, улучшенные пастища, естественные пастища, арахис, сорго, кормовое сорго, подсолнечник, пшеница и древесная растительность.

Шесть спектральных каналов из данных Landsat, а также месяц и год наблюдения использовались для предсказания типа культуры. Поскольку здесь не использовался метод временных рядов, набор данных был сокращен только до тех месяцев, о которых известно, что они имеют наземные достоверные значения (март 2011 г., сентябрь 2011 г. и февраль 2012 г.). С целью построения и оценки моделей был доступен набор данных о категориях земного покрова для 1186 участков с привязкой к земле.

Обработка данных

Наземные наблюдения земного покрова обычно проводились для областей, охватываемых несколькими пикселями изображений Landsat 7. Таким образом, 1186 наземных наблюдений категорий земного покрова расширились до 247 210 пикселей спутниковых изображений. Эти наземные пиксели считаются наблюдениями.

Доля пикселей, соответствующих каждому типу культур в наборе данных, привязанных к земле, сильно различалась. Доминирующими типами сельскохозяйственных культур были хлопок (28%), сорго (22%), естественные пастища (13%) и пшеница (11%). Девять (9) типов культур имели менее 1% от общего количества пикселей, привязанных к земле.

Для всех анализов использовалось соотношение обучения/проверки – 80/20 (т. е. для построения модели использовалась случайная выборка из 80% наземных наблюдений, а затем оценивалась прогностическая способность модели с использованием оставшихся 20% наземных данных). Количество повторений (прогонов) с разными выборками обучения/проверки варьировалось для разных методов машинного обучения и, как правило, было небольшим, поскольку большой объем данных означал, что прогоны были чрезмерно сложными для вычисления, а разница в результатах между прогонами была минимальной.

Различные вегетационные индексы (VI) рассчитывались в качестве функций данных изображений Landsat. Как описано ранее в этой главе, VI предназначены для выделения контрастов между каналами, которые могут свидетельствовать о характеристиках растительности. Функции, определяющие VI, могут быть более сложными, чем продукты и полиномиальные термины, обычно

используемые для ковариат в регрессионном анализе. Существует ряд программ для расчета VI на основе данных об отражательной способности; см., например, пакеты R «hsdar» в публикации Meyer & Lehnert (2016 г.) и «RStoolbox» у Leutner & Horning (2016 г.). Однако, учитывая очевидные внутренние несоответствия в пакете «hsdar» и требование пакета «RStoolbox» в отношении данных отражательной способности в растровом формате, выбранные индексы растительности были рассчитаны вручную с использованием формул в пакете RStoolbox и констант, предоставленных Ahmadian и др. (2016 г.) и She и др. (2015 г.).

Реализация методов

Все анализы проводились с использованием статистического программного обеспечения R. Параметры нейронной сети и бустинга деревьев регрессии были настроены вручную для получения оптимальных результатов. Нейронная сеть была подобрана с использованием пакета H2O с функцией активации «ректификатора» и скрытой компоновкой слоев: 200 узлов на уровне 1, 100 узлов на уровне 2 и 15 узлов на уровне 3. Бустинг дерева регрессии был подобран с использованием пакета «gbm». Регрессионные полиномиальные логит-регрессионные модели Ридж и Лассо были подобраны с использованием пакета R «glmnet» со 100-кратной перекрестной проверкой. Поскольку перекрестная проверка данных с чрезвычайно редкими категориями проблематична, модели регрессии были ограничены культурами с не менее, чем 10 000 наблюдений (голая почва, хлопок, кукуруза, естественные пастбища, арахис, сорго и пшеница).

Для методов регрессии, поскольку между каналами исходных изображений и индексами растительности была значительная коллинеарность, матрица плана была отфильтрована, чтобы обеспечить максимальное значение коэффициента корреляции между оставшимися ковариатами, равное 0,85. Сохраненными переменными были каналы Landsat 1, 3, 4 и 5, Нормализованный разностный вегетационный индекс (NDVI), индекс растительности с коррекцией по почве (SAVI), модифицированный нормализованный разностный водный индекс (MDWI) и индекс растительности простого отношения (SR). Все ковариаты были перецентрированы и повторно масштабированы (до среднего значения столбца, равного нулю, и до нормы столбца L1, равной единице). В модели регрессии были включены линейные парные взаимодействия и полиномиальные члены до полиномиального порядка четыре.

Для оценки результатов использовались два разных подхода, а именно матрица ошибок и ROC-кривая (рабочая характеристика приёмника). В матрице ошибок прогнозируемые классы культур сравниваются с фактическими классами. ROC показывает правильную скорость классификации бинарного классификатора при изменении порога различия.

Полученные результаты

Конфигурация нейронной сети привела к общему уровню ошибочной классификации примерно 26%, усредненному по 10 итерациям. Дисперсия для этих итераций была меньше 8,3e-06. Однако показатели неправильной классификации для определенных классов были намного хуже. Наблюдалась довольно четкая корреляция между точностью для данного класса и общей долей этого класса в выборке. Одним из возможных методов решения этой проблемы может быть использование более равномерного распределения выборки по классам, которое не является репрезентативным для истинных соотношений. В качестве альтернативы небольшая часть классов может быть исключена или объединена в один «другой» класс.

Подход с использованием бустинга дерева регрессии выявил доминирование одной переменной, канала 5 снимков Landsat, при различении типов культур. Относительное влияние этого канала по сравнению с другими пятью каналами показано ниже. Канал 5 (1,55–1,75 мкм) чувствителен к набуханию или количеству воды в растениях. В других экологических исследованиях канал 5 разделял лесные угодья, пахотные земли и водоемы. Леса оказались сравнительно более темными по тону, чем пахотные земли (светло-серые) и травяные угодья (более светлые тона). Включение производных вегетационных индексов (VI) не дало никаких улучшений. Вполне вероятно, что любая сложная взаимосвязь, захваченная предварительно обработанным VI, может быть

аналогичным образом представлена исходными переменными в анализе бустинга дерева регрессии и, следовательно, добавляет мало значения.

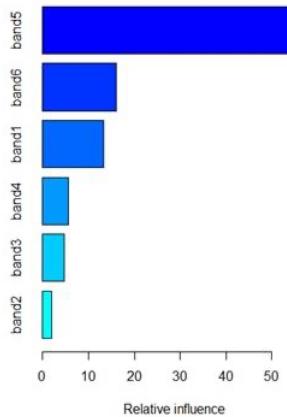


Рисунок 26: Относительное влияние каналов изображений Landsat, полученное в результате анализа бустинга дерева регрессии.

Регуляризованные модели полиномиальной логистической регрессии Лассо смогли правильно классифицировать 78% пикселей Landsat. Нелинейные и взаимодействующие члены внесли менее половины процента в общую точность классификации, но они улучшили классификацию определенных типов культур. Например, в моделях Лассо включение этих терминов обеспечило гораздо лучшую классификацию кукурузы.

Модели Ридж достигли общего уровня правильной классификации 68,5%, используя только линейные эффекты. В отличие от моделей Лассо, модели Ридж вообще не присваивали пиксели кукурузе и присвоили очень мало пикселей пшенице, а также были значительно менее точными при классификации арахиса и естественных пастбищ. Добавление нелинейных и взаимодействующих условий увеличило общую правильную классификацию до 75%, при этом существенно улучшилась классификация пшеницы, кукурузы и арахиса. В отличие от моделей Лассо, которые по своей конструкции исключают неинформативные переменные, в моделях Ридж сохраняются все ковариатные члены.

Обсуждение

Это исследование достигло своей цели по внедрению ряда методов машинного обучения в рамках анализа данных НЗ. Для всех методов наблюдались проблемы с управлением большими наборами данных и с правильной классификацией большого количества типов культур, некоторые из которых очень малы. Хотя методы легко доступны в программном обеспечении, таком как R, эти проблемы означают, что для реализации и настройки методов требуется много времени и усилий. Тем не менее, исследование подтвердило выводы других опубликованных работ о том, что эти методы имеют большие перспективы для автоматической идентификации типов сельскохозяйственных культур.

Были отмечены некоторые улучшения в реализации методов. Во-первых, в текущем анализе группы перекрестной проверки были построены путем случайного подмножества полного набора данных с использованием метода по умолчанию, поставляемого в пакете «glmnet». Лучшей стратегией было бы использование набора групп, построенных таким образом, чтобы каждая группа содержала примерно такую же долю наблюдений различных категорий ответов, которая присутствует в полных данных. Во-вторых, проблему выявления редких ответов можно было бы лучше решить путем искусственного завышения количества наблюдений в этих категориях с использованием методов начальной загрузки или подвыборки или путем применения различных степеней сокращения коэффициентов для различных категорий ответов. В-третьих, дополнительная диагностика модели может помочь лучше понять уровень ошибочной классификации для

различных категорий сельскохозяйственных культур и, следовательно, понизить его. В-четвертых, можно рассмотреть дополнительные ковариаты, такие как производные переменные MODIS и индексы Tasseled Cap. В-пятых, и последнее, другие методы регуляризации для регрессионных моделей, такие как эластичная сеть, могут быть оценены с учетом различной вычислительной нагрузки.

4.5 ПРИМЕР 2: Прогноз урожайности (США)

Это исследование, о котором писали Gao и др. (2017 г.), было направлено на определение стадии роста сельскохозяйственных культур и прогнозирование урожайности на основе данных дистанционного зондирования, чтобы способствовать более эффективному управлению орошением, повышению продовольственной безопасности и эффективной торговле товарами.

Данные НЗ были предназначены для дополнения наземных наблюдений за развитием и состоянием урожая, которые в настоящее время публикуются USDA NASS. В рамках этой системы предоставляются визуальные наблюдения и субъективные оценки урожая, которые еженедельно публикуются в отчетах NASS Crop Progress.

Gao и др. расширяют применение слияния данных до масштабов подполей с целью фенологического картирования сельскохозяйственных ландшафтов. Основная цель состояла в том, чтобы проверить способность этой методологии слияния данных Landsat-MODIS отображать фенологию культур с разрешением 30 м, где можно различать стадии роста различных культур. Карты развития посевов были составлены на интенсивно возделываемых территориях в центральной части Айовы.

Характеристики данных

Область исследования находилась в центральной части Айовы, в зоне неорошающего земледелия в районе Кукурузного пояса США. Кукуруза и соя являются тут основными культурами.

Данные Landsat и MODIS за девять лет с 2001 по 2014 год были выбраны для создания объединенных временных рядов отражательной способности Landsat-MODIS с целью картирования фенологии сельскохозяйственных культур.

Другие источники данных, включенные в исследование, кратко описаны здесь:

- *Слой данных о возделываемых землях (CDL).* Данные NASS CDL с 30-метровым разрешением за период с 2001 по 2014 год были загружены и использованы для анализа фенологических моделей основных типов культур на изучаемой территории.
- *Данные измерительных вышек.* Метеорологические данные и локальные измерения поверхностного потока были собраны на паре прилегающих сельскохозяйственных полей к северу от Эймса (недалеко от Бакай), штат Айова, в округе Хардин.
- *Полевые биофизические данные.* Полевые наблюдения за развитием посевов проводились в период вегетации с 2010 по 2013 год на двух участках наблюдения в округе Даллас, штат Айова, расположенных в пределах исследуемой территории.

Обработка данных

Анализ, проведенный в этом исследовании, включает два основных компонента: слияние данных и извлечение фенологии. Вся схема обработки показана ниже. Для получения дополнительной информации см. Gao и др. (2017 г.).



Блок-схема обработки и анализа данных по фенологии сельскохозяйственных культур. Прямоугольники обозначают продукты данных, а эллипсы – этапы/модули обработки.

Источник: Gao и др. (2017 г.). На пути к картированию прогресса сельскохозяйственных культур в полевых масштабах путем слияния изображений Landsat и MODIS. <http://www.sciencedirect.com/science/article/pii/S0034425716304369>

Выбор данных

Только самые четкие сцены Landsat использовались в качестве изображений с высоким разрешением для сопряжения с данными MODIS в качестве входных данных для STARFM. Частично четкие изображения Landsat не использовались в процессе слияния данных. Изображения Landsat 7, у которых произошел сбой SLC после мая 2003 г., не использовались в качестве входных парных изображений, поскольку в объединенных изображениях останутся пропуски SLC-off, что повлияет на последующее фенологическое картирование. Однако все действительные и четкие пиксели (как указано в маске облаков) со всех доступных изображений Landsat, включая частично четкие изображения и изображения Landsat 7 SLC-off, были использованы в самом фенологическом картировании.

Метод

Подход STARFM использовался для получения многолетних ежедневных 30-метровых показателей отражения поверхности и временных рядов NDVI, которые фиксировали пространственную и сезонную изменчивость между классами кукурузы, сои и лесной растительности, которые характеризуют этот регион. Подход к слиянию данных STARFM был подробно описан в публикации Gao и др. (2006 г.).

С целью извлечения фенологии урожая из ежедневного временного ряда NDVI, были применены два подхода – MODIS и TIMESAT. Подход TIMESAT опирается на предварительно определенный порог (процентное изменение сезонной амплитуды между минимальным и максимальным значениями NDVI) для определения начала сезона, конца сезона и продолжительности сезона. Фенологический подход MODIS использует анализ кривизны для определения точек перегиба на аппроксимирующей кривой. Двойная логистическая функция была выбрана для соответствия временному ряду NDVI для обоих подходов.

Показатели фенологии сельскохозяйственных культур, полученные из объединенных данных Landsat-MODIS, сравнивались со стадиями роста сельскохозяйственных культур,

зарегистрированными на уровне районов и округов из отчетов NASS CP. Наблюдения в масштабе поля исследовались с использованием соответствующих пикселей Landsat.

На рисунке ниже показаны аппроксимированные кривые NDVI за 2011 г. для пикселей кукурузы и сои на двух участках измерительной вышки в Саут-Форк, наложенные на этапы роста урожая из отчетов CP NASS на уровне округа. Также показаны ежедневные оценки NEE и GPP, полученные на основе измерений ковариации вихрей в течение вегетационного периода в 2011 г. Общее соответствие во времени кривых NDVI и поглощения углерода растениями показывает, что временной ряд NDVI ведет себя разумно на этих двух полях во взаимосвязи с этапами CP NASS.

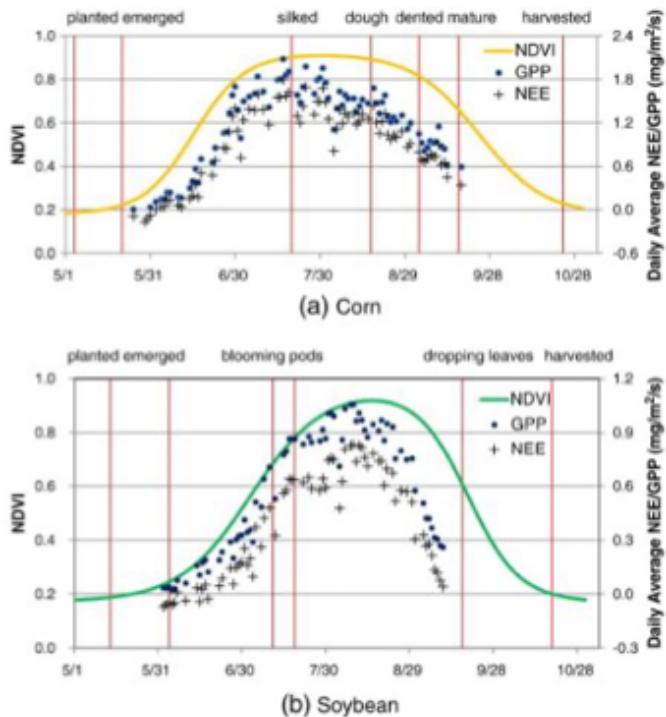


Рисунок 27: Кривые NDVI для пикселей кукурузы (а) и сои (б) по сравнению с ежедневными NEE и GPP (вторичная ось, измеряемая только во время вегетационного периода) и стадиями роста урожая из отчетов NASS на уровне округов в 2011 г. (красные вертикальные линии). Можно выявить четкие соответствия между временными рядами NDVI, NEE и GPP во время вегетативных стадий сельскохозяйственных культур при разрешении пикселей Landsat.

Источник: Gao и др. (2017 г.). На пути к картированию прогресса сельскохозяйственных культур в полевых масштабах путем слияния изображений Landsat и MODIS. <http://www.sciencedirect.com/science/article/pii/S0034425716304369>

Проблемы с качеством

Применение этих методов в режиме реального времени – еще одна область, требующая дополнительных исследований. С оперативной точки зрения сведения об этапах и условиях роста сельскохозяйственных культур необходимо сообщать еженедельно в течение текущего вегетационного периода. Программы фенологии, используемые в этой статье, не были предназначены для картирования внутрисезонной фенологии культур с использованием данных за неполные годы. Однако подход, основанный на модели роста, представленный Sakamoto и др., 2010 г. и Sakamoto и др., 2011 г., может быть адаптирован для полевого применения в реальном времени с использованием объединенных временных рядов Landsat-MODIS.

Полученные результаты

Подход STARFM использовался для получения многолетних ежедневных 30-метровых показателей отражения поверхности и временных рядов NDVI, которые фиксировали пространственную и сезонную изменчивость между классами кукурузы, сои и лесной растительности, которые характеризуют этот регион. Погрешности между совмещенным и наблюдаемым Нормализованным разностным вегетационным индексом (NDVI) составляли от -0,011 до 0,028, а средние абсолютные различия составляли <0,083 за все годы с 2001 по 2014 гг. Эти отклонения частично объясняются разницей в углах восприятия датчиков, волн и обзора Landsat и MODIS, и было продемонстрировано, что важно характеризовать эти отклонения, чтобы полностью понять производительность синтеза.

Используя временные ряды карт NDVI с разрешением 30 м, полученных в результате слияния данных Landsat-MODIS, фенологические показатели урожая были извлечены с использованием двух установленных фенологических программ (MODIS и TIMESAT). В то время как оба подхода генерировали сопоставимые показатели, подход кривизны MODIS был сочен наиболее практичным, поскольку он не требует указания предварительно определенных пороговых значений, как того требует подход TIMESAT.

Результаты исследования позволяют предположить, что существует возможность надежного ретроспективного прогнозирования на основе данных дистанционного зондирования, например, в качестве исходных данных для имитационных моделей урожая. Это исследование продемонстрировало, что данные временных рядов из нескольких источников дистанционного зондирования могут быть эффективно интегрированы и использованы для определения основных физиологических стадий сельскохозяйственных культур в полевых масштабах. При разрешении 30 метров фенология посевов кукурузы и сои может быть четко разделена и количественно определена для размеров полей, типичных для США. Эта возможность находит применение в управлении сельскохозяйственными угодьями и прогнозировании урожайности. Однако регулярное картирование стадий роста сельскохозяйственных культур в течение сезона по-прежнему сопряжено с трудностями в силу того, что для картирования в режиме, близком к реальному времени, данные имеются лишь за неполный год. Использование в реальном времени потребует соответствующего фенологического подхода и частого получения изображений с помощью дистанционного зондирования при ясном небе. Слияние данных спутниковых изображений может сыграть важную роль в достижении этой цели.

4.6 ПРИМЕР 3: Мониторинг развития сельскохозяйственных культур (Южная Корея)

Цель этого исследования заключалась в изучении применимости спутниковых данных геостационарного формирователя цветных изображений океана (GOCI) с высоким времененным разрешением для мониторинга процесса выращивания риса-сырца в Южной Корее. Меры мониторинга включали в себя спектральный анализ и профили вегетационного индекса. 27 июня 2010 года было успешно начато осуществление программы GOCI. Работающие с этой программой спутники предназначены для обнаружения, мониторинга и прогнозирования региональных океанических явлений вокруг Кореи и оснащены восемью спектральными каналами (шесть видимых, два – инфракрасных).

Характеристики данных

В этом исследовании были выбраны два района риса-сырца; один был расположен в Kyehwa и соответствует пикслю GOCI с координатами $35^{\circ}46'37N$ и $126^{\circ}41'03E$ (рис. 1b). Другой был в Kimjae и соответствует пикслю GOCI с координатами $35^{\circ}44'59N$ и $126^{\circ}52'15E$ (рис. 1c).

Участок в Kimjae представляет собой двойной урожай ячменя и раннеспелых сортов риса, а участок в Kyehwa представляет собой наиболее популярное выращивание риса-сырца среднего позднеспелого сорта. Поскольку эти исследуемые участки относительно однородны, несмотря на небольшие рисовые поля, временная динамика различных культур должна быть распознаваема при ежедневном анализе данных спутниковых изображений.

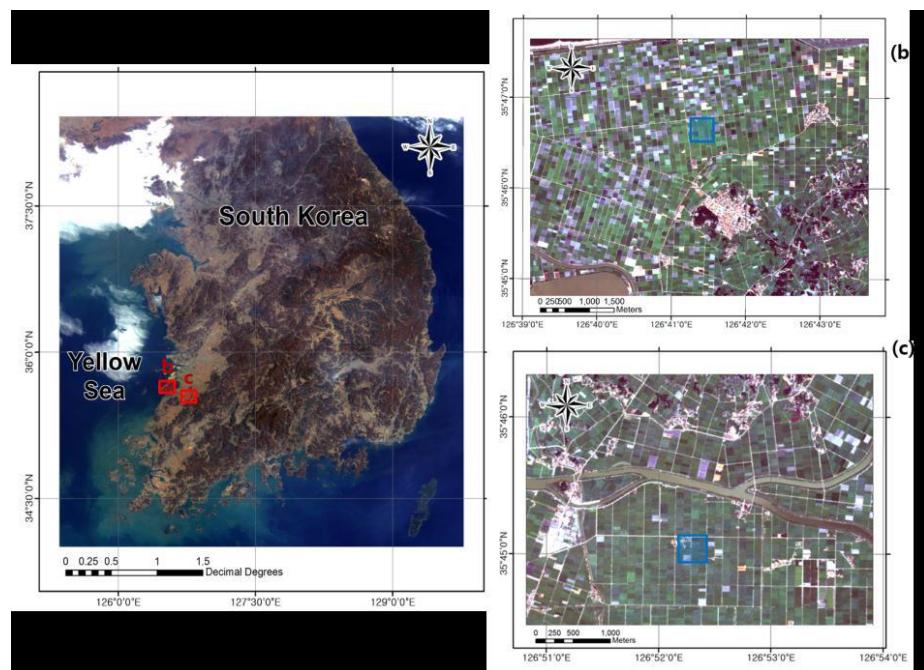


Рисунок 28: Район исследования. (а) Композитное изображение красного, зеленого и синего цветов (RGB), полученное 1 апреля 2011 г. с помощью геостационарного формирователя цветных изображений океана (GOCI). Красные прямоугольники (б) и (с) на рисунке (а) показаны отдельно на рисунках (б) и (с), соответственно. Они представляют подробные изображения с использованием мультиспектральных данных RapidEye высокого разрешения, полученных 5 августа 2011 г. для рисунка (б) и 11 октября 2011 г. – для (с). Синие прямоугольники на (б) и (с) геометрически совпадают с соответствующими пикселями спутникового наблюдения.

Источник: Yeom, J-M., & Kim, H-O. (2015 г.), Сравнение NDVI из данных GOCI и MODIS для улучшения оценки временной динамики урожая в случае риса-сырца. Дистанционное зондирование (7), 11326 – 11343.

Наземные измерения проводились с использованием мультиспектрального радиометра (MSR) с целью оценки спутниковых профилей растительности для сравнительного анализа. Полевые измерения проводились с июня по октябрь 2014 года. Чтобы получить характеристики развития урожая риса-сырца, измерения проводились в восемь различных дат на основе графика выращивания, включая пересадку и сбор урожая.

Обработка данных

Для спутниковых изображений GOCI была выполнена дальнейшая предварительная обработка, включая преобразование цифровых чисел (DN) в излучение, маскирование облаков и атмосферную коррекцию для расчета коэффициента отражения поверхности. Подробнее см. Yeom & Kim (2015 г.).

Выбор данных

В исследовании сравнивались два набора данных оптических спутников наблюдения Земли с одинаковым пространственным разрешением 500 м; один набор был с геостационарного (GOCI), а другой – с солнечно-синхронного спутника (MODIS). Пятый и восьмой каналы GOCI использовались для расчета Нормализованного разностного вегетационного индекса (NDVI). Для сравнения в качестве эталона использовались продукты MODIS NDVI. В этом исследовании проанализированы данные за 2011-2014 годы.

Проблемы с качеством

Для сравнения с NDVI, рассчитанного по спутниковым данным среднего разрешения, с натурными измерениями, необходимо рассмотреть проблему смешанных пикселей. Поскольку рисовые поля в Южной Корее относительно очень малы, очень трудно наблюдать несмешанное спектральное значение для рисовых полей по данным GOCI со средним пространственным разрешением.

Метод

Вегетационный индекс был рассчитан из данных GOCI на основе Функции распределения двунаправленной отражательной способности (BRDF) с поправкой на отражательную способность, которая затем использовалась для визуального анализа сезонной динамики урожая. Затем эти вегетационные индексы сравнивали с индексами, рассчитанными с использованием сканирующего спектрорадиометра среднего разрешения (MODIS) – Нормализованного разностного вегетационного индекса (NDVI), основанного на коэффициенте отражения в nadir, скорректированного BRDF.

Полученные результаты

Результаты показывают явные преимущества метода GOCI, который обеспечивает в четыре раза лучшее временное разрешение, чем комбинированные датчики MODIS, интерпретируя едва различимые характеристики развития растительности. В целом, GOCI и MODIS продемонстрировали аналогичные временные колебания NDVI при благоприятных погодных условиях, потому что они могут обеспечить достаточное количество безоблачных наблюдений для моделирования полной инверсии BRDF. Однако во время сезона дождей, с его длительными периодами ливней и большим количеством облачных дней, GOCI оказался более полезным для выделения безоблачных или менее облачных областей путем объединения восьми изображений для расчета репрезентативных ежедневных данных.

Исследование также показало, что GOCI BAR NDVI более полезен для мониторинга сигнала урожая, чем широко используемый MVC NDVI. Авторы пришли к выводу, что очень высокое временное разрешение GOCI весьма полезно для мониторинга развития сельскохозяйственных культур и может предоставить улучшенную информацию по фенологии.

4.7. ПРИМЕР 4: Операционная основа для картирования крупных площадей прошлого и настоящего земледелия с использованием сезонных изображений Landsat и метрики временных рядов (Австралия)

В Квинсленде, Австралия, Департамент науки, информационных технологий и инноваций (DSITI) уже много лет использует спутниковые снимки для статистического анализа. Хотя DSITI формально не входит в состав Целевой группы, его исследование включено в эту главу в качестве тематического исследования для НСУ.

Управление земельными ресурсами с целью удовлетворения потребностей населения является обязательным условием, общим для всех уровней правительства в Австралии, и в этом исследовании, проведенном DSITI, используются данные спутниковых снимков для выявления активно растущих сельскохозяйственных культур в Квинсленде, Австралия.

Для крупных районов широкомасштабного земледелия в Квинсленде, Австралия, полный архив временных рядов Landsat (LTS) с 1987 по 2015 год использовался в мультивременном картографическом подходе, при котором пространственная, спектральная и временная информация объединялась в процессе моделирования урожая, поддерживаемого данными обучения, отобранными в пространстве и времени для классов Crop и No-Crop.

Информация о временных интервалах в летний и зимний вегетационные периоды обобщалась за каждый год в сочетании с различными индексами растительности и коэффициентами каналов, рассчитанными на основе синтетического спектрального изображения на основе пикселей в середине сезона. Вся доступная временная информация была пространственно агрегирована в масштабе сегментов изображения в синтетическом изображении в середине сезона для каждого вегетационного периода и использовалась для обучения ряда различных прогностических моделей с целью классификации Crop и No-Crop.

Полученные результаты свидетельствуют о том, что при наличии соответствующих данных о подготовке кадров и подтверждении достоверности данных, подход, разработанный и применяемый в настоящем исследовании, подходит для картирования крупных площадей с использованием архива Landsat. Этот подход можно было бы перенести на другие спутниковые изображения с достаточным пространственным, спектральным и временным разрешением для получения пояснительных переменных, используемых в настоящем исследовании.

Характеристики данных

Областью исследования была Западная зона возделывания сельскохозяйственных культур (WCZ) в штате Квинсленд, Австралия (см. Рисунок 29).

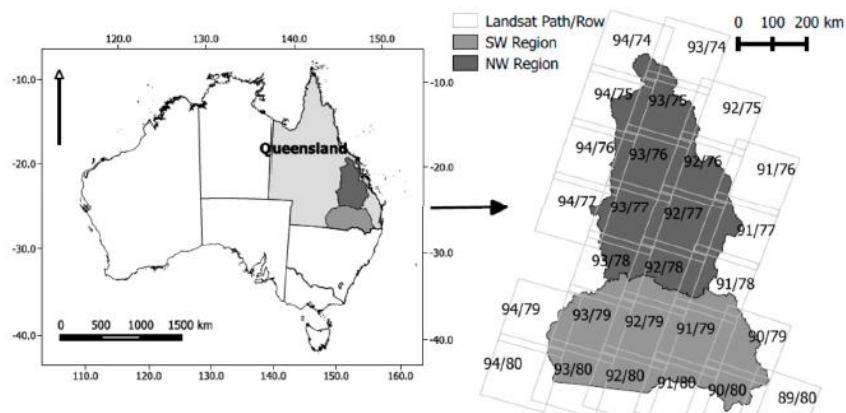


Рисунок 29: Региональное подразделение WCZ Квинсленда на северный и южный регионы: «СЗ» и «ЮЗ». Накладывается покрытие данных Landsat, основанное на World Reference System 2 (WRS2).

Источник: Schmidt, M., Pringle, M., Rakshesh, D., Denham, R. и D. Tindall. 2016 г. Структура картирования крупных площадей прошлой и настоящей сельскохозяйственной деятельности с использованием сезонных изображений Landsat и показателей временных рядов (<http://www.mdpi.com/2072-4292/8/4/312>).

Были использованы все доступные данные Landsat 5 TM, Landsat 7 ETM+ и Landsat 8 OLI по 27 зонам охвата с облачностью менее 60%. Этот критерий был выбран для снижения потенциального шума в данных из-за несовершенной маскировки облаков/теней облаков и вариаций водяного пара.

Все изображения были замаскированы облаками с помощью региональной адаптированной версии Fmask²², а затем скорректированы с помощью NBAR²³; этот последний шаг преобразует единицы изображения в отражательную способность нижней части атмосферы, масштабированную до интервала (0, 1). Высокая внутренняя геометрическая согласованность изображений Landsat на основе Геологической службы США позволяет проводить дальнейшую обработку и анализ данных без дополнительной геометрической корректировки.

Обработка данных

Все архивные изображения использовались для создания синтетического спектрального изображения для каждого сезона и года. Эти изображения были сгенерированы для каждого пикселя и каждого спектрального канала, исключая тепловой канал, основанный на статистической модели. На рис. 30 представлен пример, призванный продемонстрировать для одного пикселя, как значения синтетической отражательной способности Landsat были получены для середины вегетационного периода, t_0 , в течение определенного года. Выбросы были оценены, а затем 4 ближайших соседа к t_0 были использованы в линейной регрессии для прогнозирования отражательной способности в t_0 . Для получения дополнительной информации см. Schmidt и др. (2016 г.).

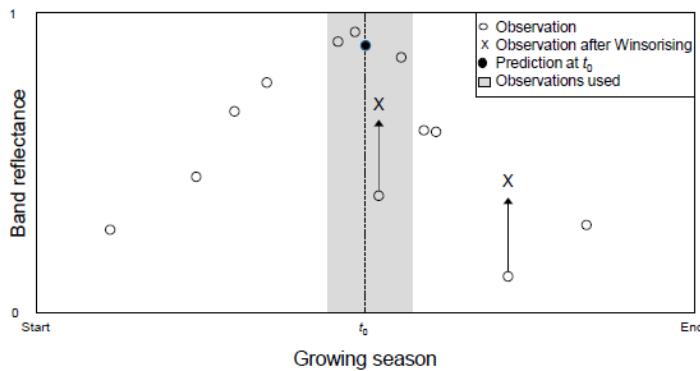


Рисунок 30: Пример процесса создания синтетического значения коэффициента отражения.

Источник: Schmidt, M., Pringle, M., Rakshesh, D., Denham, R. и D. Tindall. 2016 г. Структура картирования крупных площадей прошлой и настоящей сельскохозяйственной деятельности с использованием сезонных изображений Landsat и показателей временных рядов (<http://www.mdpi.com/2072-4292/8/4/312>).

Самая последняя Программа картирования землепользования Квинсленда (QLUMP) использовалась для маскирования областей из каждого синтетического изображения, которые были связаны с землепользованием, не представляющим интереса для этого исследования, таким как

²² Flood, N. Непрерывность данных об отражательной способности между Landsat-7 ETM+ и Landsat-8 OLI, как для верхней части атмосферы, так и для поверхностной отражательной способности: исследование австралийского ландшафта. *Remote Sens.* 2014, 6, 7952–7970.

Frantz, D.; Roder, A.; Udelhoven, T.; Schmidt, M. Повышение возможности обнаружения облаков и их теней на многовременных снимках Landsat в засушливых районах: расширение Fmask. *IEEE GeoSci. Письмо о дистанционных датчиках.* 2015, 12, 1242–1246.

Zhu, Z.; Woodcock, C.E. Обнаружение облаков и их теней на основе объектов на снимках Landsat. *Дистанционный датчик окружающей среды.* 2012, 118, 83–94.

²³ Flood, N. Непрерывность данных об отражательной способности между Landsat-7 ETM+ и Landsat-8 OLI, как для верхней части атмосферы, так и для поверхностной отражательной способности: исследование австралийского ландшафта. *Remote Sens.* 2014, 6, 7952–7970.

заповедные зоны, лесное хозяйство и интенсивное использование. Также для каждого синтетического изображения маскировались водоемы и дороги.

Чтобы отразить тот факт, что земля управляет отдельными участками, синтетические изображения были сегментированы на однородные области с использованием программного пакета с открытым исходным кодом RSGISLib с привязкой к Python. Сегментированное изображение было сохранено в поддерживаемом GDAL формате изображения KEA, где сегменты изображения связаны с таблицей растровых атрибутов (RAT) и используются в качестве основы для всего дальнейшего анализа. Этот шаг представляет собой метод сокращения данных путем перехода от анализа на основе пикселей к пространственным медианам на сегмент для всех переменных классификации. На заключительном этапе использовалась цифровая модель рельефа высотой 1-s для маскировки сегментов с уклоном более 10%, поскольку земледелие, как правило, не проводится на крутых склонах в пределах исследуемого региона.

Подмножество синтетического изображения в середине сезона для зимы 2014 года представлено на рисунке 31.

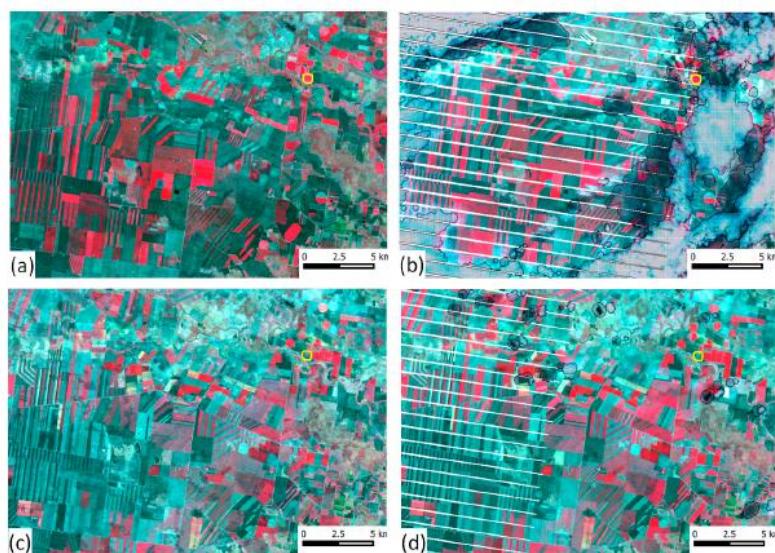


Рисунок 31: Подмножество синтетического изображения в середине сезона (R/G/B с каналами Landsat TM 4/3/2) для зимы 2014 г. (a) сегмент 90/79 и ближайшее изображение Landsat (ETM+) (b) к межсезонному интервалу для зимы 20 сентября 2014 г. На сегменте (c) показано синтетическое изображение середины сезона летом 2014 г.; и (d) ближайший снимок Landsat от 8 февраля (ETM+). Облака F-маски и маски теней от облаков показаны черной штриховкой, желтым кружком на (a) отмечено центральное круговое поле орошения.

Источник: Schmidt, M., Pringle, M., Rakheesh, D., Denham, R. и D. Tindall. 2016 г. Структура картирования крупных площадей прошлой и настоящей сельскохозяйственной деятельности с использованием сезонных изображений Landsat и показателей временных рядов (<http://www.mdpi.com/2072-4292/8/4/312>).

Переменные классификации

Был рассчитан ряд переменных, чтобы зафиксировать признаки отражения и временные характеристики, которые использовались для классификации сегментов изображения (таблица 11).

Таблица 11: Переменные дистанционного зондирования, рассчитанные для каждого пикселя, которые были пространственно агрегированы и добавлены в таблицу растровых атрибутов сегментированных синтетических изображений.

Индексы растительности, временные переменные и соотношения каналов

- 1** Нормализованный разностный вегетационный индекс (NDVI)
- 2** Модифицированный индекс поглощения хлорофилла (MCARI)
- 3** Перенормализованный разностный индекс вегетации (RDVI)
- 4** Триангулярный вегетационный индекс (TVI)
- 5** Модифицированное простое соотношение (MSR)
- 6** Нормализованный разностный коэффициент выгорания (NDBR)
- 7** Сезонная дисперсия NDVI (ndviTsVr)
- 8** Сезонный минимум NDVI (ndviTsMn)
- 9** Сезонный максимум NDVI (ndviTsMx)
- 10** Коэффициент сезонной вариации NDVI (ndviTsCV)
- 11** Сезонные колебания NDVI (ndviTsRng)
- 12** Градиент NDVI вверх (от первого минимума до максимума) (ndviTsGr1)
- 13** Градиент NDVI вниз (от максимума до второго минимума) (ndviTsGr2)
- 14** Максимум временных рядов NDVI (ndviTsDyMx)
- 15** $b7_b3/(b7 + b3)$ (nr73)
- 16** $b7_b2/(b7 + b2)$ (nr72)
- 17** $b5_b7/(b5 + b7)$ (nr57)
- 18** $b4_b5/(b4 + b5)$ (nr45)
- 19** $b5_b3/(b5 + b3)$ (nr53)
- 20** $b5_b2/(b5 + b2)$ (nr52)
- 21** $b4_b2/(b4 + b2)$ (nr42)
- 22** $b2/b3$ (r23)
- 23** $b4/b3$ (r43)

Источник: Schmidt, M., Pringle, M., Rakshesh, D., Denham, R. и D. Tindall. 2016 г. Структура картирования крупных площадей прошлой и настоящей сельскохозяйственной деятельности с использованием сезонных изображений Landsat и показателей временных рядов (<http://www.mdpi.com/2072-4292/8/4/312>).

В исследовании использовались полевые обучающие данные, собранные в период с сентября 1999 г. по февраль 2012 г. Все записи были перекодированы как Сгроп или Но-Сгроп. Дополнительный набор записей был создан путем компьютерной интерпретации сезонных синтетических изображений, отобранных случайным образом в вегетационные периоды за разные годы.

К одним и тем же обучающим данным применялись четыре алгоритма классификации; SVM, полиномиальная логистическая регрессия, классификатор дерева решений и RF, все они реализованы с помощью программного обеспечения R. Проверка показала, что точность прогнозирования варьируется в зависимости от вегетационного периода и региона, а лучше всего работает случайный классификатор леса, где $\kappa = 0,88\text{--}0,91$ для летнего вегетационного периода и $\kappa = 0,91\text{--}0,97$ для зимнего вегетационного периода, и, таким образом, он подходит для картирования как текущей, так и исторической сельскохозяйственной деятельности.

Выбор данных

Были использованы все доступные данные Landsat 5 TM, Landsat 7 ETM+ и Landsat 8 OLI с 1987 г. по 27 зонам охвата с облачностью менее 60%. Для большинства сезонов есть 10 или более изображений Landsat с облачностью менее 60%. Этот порог был применен для снижения риска появления артефактов на изображениях из-за несовершенного маскирования облаков и их теней, тонких необнаруженных облаков и вариаций водяного пара, которые могут влиять на значения отражательной способности отдельного пикселя. Высокий уровень облачности также повлиял на геометрическую точность данных Landsat 5. Данные соответствовали цели исследования.

Проблемы с качеством

Относительно незначительная проблема заключается в том, что сегменты изображения для обрезанного поля не всегда могут быть четкими на границах поля, поэтому некоторые

несоответствия с истинными контурами поля очевидны. Кроме того, сравнение синтетического изображения середины сезона эффективно, если культуры озеленяются в определенное время, но, если культура высаживается поздно, она может не выглядеть зеленой на синтетическом изображении. В некоторых случаях это может привести к неправильной классификации сегментов как неурожайных.

Недавно авторы использовали аналогичную структуру и расширили картографический подход к широким типам сельскохозяйственных культур (Pringle и др., 2018 г.). Этот подход включает в себя все доступные изображения Landsat, начиная с 1987 года и до настоящего времени, а также дополнительно включает изображения Sentinel-2 (начиная с 2015 года). В случае больших временных промежутков в качестве резервного варианта также используются изображения MODIS. В этом обновленном подходе в качестве входных данных для GEOBIA использовалось временное адаптивное сезонное составное изображение во время максимального EVI, за которым следовал древовидный алгоритм классификации изображений фенологической метрики. Типы культур (i) зерновых (пшеница, ячмень, овес) и (ii) зернобобовых (нут) на зиму, а также (i) крупнозерновых и бобовых (сорго, кукуруза, маш, соя), и (ii) летний урожай хлопка сравнивался со значениями официальной статистики (Рисунок 32).

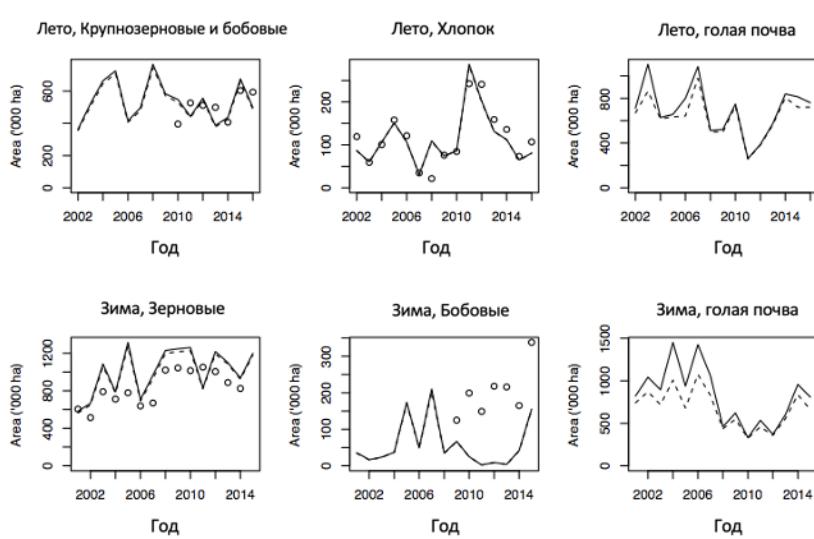


Рисунок 32: Площадь, отведенная различным группам в каждый вегетационный период, согласно прогнозу модели DSITI (линии). Сплошная линия представляет собой итоговую величину для всего исследуемого региона; пунктирная линия представляет собой общую величину, связанную с земледелием.

Источник: Pringle, M., Schmidt, M. и Tindall, D. (2018 г.): Моделирование временных рядов с использованием нескольких сезонов и нескольких датчиков на основе геостатистических концепций для прогнозирования широких групп сельскохозяйственных культур.

Для летнего вегетационного периода наблюдается превосходное соответствие между средними по штату картами типов культур и официальными данными о товарах (ABARES, 2016 г.). Переоценка класса «Зерновые» похожа на недооценку класса «Плод бобовых». Авторы обнаружили небольшую разницу между суммами, связанными с сельскохозяйственными культурами, для их исследуемого региона.

4.8 ПРИМЕР 5: Оперативный подход к интегрированному моделированию урожайности с использованием дистанционного зондирования, агроклиматических данных и данных обследований (Канада)

Статистическое управление Канады – это НСУ, которое использует данные дистанционного зондирования с начала 1980-х годов для таких областей, как перепись и проверка результатов обследований, картирование сельскохозяйственных культур и оценка площадей, а также разработка вспомогательных программ. В 2016 году Статистическое управление Канады стало первым НСУ, заменившим статистическое обследование подходом, основанным на модели дистанционного зондирования. Целью модели было произвести предварительную оценку ожидаемой урожайности сельскохозяйственных культур в конце лета с использованием информации из существующих источников данных.

Хотя Статистическое управление Канады не разработало пилотный проект в рамках Целевой группы, их тематическое исследование отмечено, как пример НСУ, использующего данные НЗ для производства официальной статистики на практике.

Статистическое управление Канады исследовало использование данных дистанционного зондирования, агроклиматических данных и данных обследований для моделирования надежных оценок урожайности сельскохозяйственных культур в качестве предварительной оценки ноябрьского обследования фермерских хозяйств по случаю выпуска своей серии отчетов об урожае.

Первоначально эти оценки были доступны до того, как были опубликованы оценки сентябрьского обследования фермерских хозяйств. Однако с 2016 года обследование фермерских хозяйств не проводится, и публикуются только оценки, основанные на модели. Процесс до 2016 года описан в этом тематическом исследовании, чтобы продемонстрировать порядок действий, который Статистическое управление Канады использовало для замены оценок, полученных в результате обследования, оценками, основанными на модели. Работа была завершена Секцией дистанционного зондирования и геопространственного анализа Отдела сельского хозяйства и Отделом методов обследования предприятий Статистического управления Канады в сотрудничестве с Министерством сельского хозяйства и продовольствия Канады (AAFC).

Характеристики данных

Методология моделирования урожайности была разработана и протестирована на культурах, которые обычно публикуются на уровне провинций и на национальном уровне в сентябрьском обследовании фермерских хозяйств, как показано в таблице 12. На пять перечисленных провинций приходится примерно 98% сельскохозяйственных земель Канады, в различных климатических зонах и с различными типами почв. Культуры, на долю которых приходится примерно 85% доходов из 19 перечисленных культур, называются семью основными культурами.

Таблица 12: Культуры, обычно публикуемые в результатах сентябрьского обследования фермерских хозяйств, по провинциям. Информация сгруппирована по типу урожая (отображается в виде заголовков строк), по провинциям (отображается в виде заголовков столбцов).					
Тип культуры	Провинция				
	Квебек	Онтарио	Манитоба	Саскачеван	Альберта
7 основных культур					
Ячмень	X	X	X	X	X
Канола	X	X	X	X	X
Кукуруза на зерно	X	X	X		
Твердые сорта пшеницы				X	X
Овес	X	X	X	X	X
Соевые бобы	X	X	X		
Яровая пшеница	X	X	X	X	X
12 дополнительных культур					
Канареечное семя				X	
Нут				X	X
Цветные бобы		X	X		

Озимая рожь		X	X	X	X
Горох кормовой			X	X	X
Льняное семя			X	X	X
Чечевица				X	
Смесь из зерен	X	X	X	X	X
Горчичное зерно				X	X
Семена подсолнуха			X		
Белая фасоль		X	X		
Озимая пшеница	X	X	X	X	X

Примечание. Кормовая кукуруза обычно публикуется в сентябрьском обзоре фермерских хозяйств. Однако она не была смоделирована из-за отсутствия оценок урожайности в июльском обзоре фермерских хозяйств.

Источник: Статистическое управление Канады (2015 г.). Интегрированное моделирование урожайности с использованием дистанционного зондирования, агроклиматических данных и данных обследований.
http://www23.statcan.gc.ca/imdb-bmdi/document/5225_D1_T9_V1-eng.htm

В методологии моделирования использовано три источника данных: 1) спутниковые данные с низким разрешением, используемые в рамках Программы оценки состояния сельскохозяйственных культур Статистического управления Канады; 2) данные серии отчетов о сельскохозяйственных культурах Статистического управления Канады и 3) агроклиматические данные по сельскохозяйственным регионам Канады.

Обработка данных

Нормализованный разностный вегетационный индекс (NDVI)

Спектральный индекс вегетации, NDVI, использовался в качестве заменителя фотосинтетического потенциала. NDVI представляет собой нормализованное отношение коэффициентов отражения в ближнем инфракрасном (NIR) и красном (R) диапазонах ($NDVI = (\rho_{NIR} - \rho_R) / (\rho_{NIR} + \rho_R)$) и варьируется от -1 до 1 со значениями, близкими к единице, что указывает на высокое содержание растительности, а значения, близкие к нулю, указывают на отсутствие растительности на голой земле.

Данные NDVI обрабатывались на постоянной основе в течение сельскохозяйственного вегетационного периода (с апреля по октябрь) для всей территории Канады. Статистическое управление Канады располагает временными рядами данных NDVI с 1987 г. по настоящее время, которые включают годы сильной засухи и рекордного с/х производства. Ежедневные изображения NDVI были преобразованы в семидневные композиты, как было описано Latifovic и др., (2005 г.), и Статистическое управление Канады усовершенствовало эту методологию, чтобы свести к минимуму или устранить ошибки NDVI, вызванные наличием облаков (Bédard, 2010 г.).

Статистические данные NDVI возделываемых земель по сельскохозяйственным регионам переписи (CAR) рассчитывались и сохранялись в реляционной базе данных для каждой недельной составной части NDVI. С целью получения среднего значения NDVI для пахотных земель в каждом из сельскохозяйственных регионов переписи были извлечены только элементы изображения NDVI или пиксели, которые географически совпадают с базой данных сельскохозяйственного земельного покрова, созданной AAFC в рамках ежегодной инвентаризации сельскохозяйственных культур. Файл сельскохозяйственного земельного покрова и связанный с ним файл метаданных, подготовленный AAFC, доступны по ссылке www.geobase.ca/geobase/en/data/landcover/index.html.

После вычисления средних значений NDVI они были импортированы в качестве одной из баз данных входных переменных в модели сельскохозяйственных культур в виде трехнедельных скользящих средних с 18-й по 36-ю неделю (с мая по август).

Площадь съемки и данные об урожайности

Еще одним набором данных, использованным в модели, были обследования серии отчетов о полевых культурах Статистического управления Канады. Эти обследования позволяют получить информацию о зерновых и других полевых культурах, хранящихся на фермах (мартовские,

июльские, сентябрьские и декабрьские обследования хозяйств), посевных площадях (мартовские, июньские, июльские, сентябрьские и ноябрьские обследования хозяйств), убранных площадях, ожидаемой урожайности и полевой продукции сельскохозяйственных культур (обследования ферм в июле, сентябре и ноябре). Эти данные обеспечивают точную и своевременную оценку планов посева, засеянных и убранных площадей, производства, урожайности и сельскохозяйственных запасов основных полевых культур в Канаде на уровне провинций.

Что касается обильных культур, то в качестве исходных переменных для моделей использовались оценки урожайности на уровне CAR, полученные в ходе июльских и ноябрьских обследований фермерских хозяйств с 1987 года по настоящее время, а оценки урожайности на основе сентябрьского обследования фермерских хозяйств и ноябряского обследования фермерских хозяйств использовались для проверки точности результатов моделирования урожайности. По менее распространенным культурам данные обследования собирались на уровне провинций.

Агроклиматические показатели

Климатические данные, собранные в течение вегетационного периода, были третьим источником данных, используемым для моделирования урожайности сельскохозяйственных культур. Данные о суточной температуре и осадках на станциях, предоставленные Министерством окружающей среды Канады и другими учреждениями-партнерами, использовались для составления прогнозов на основе климата (Chipanshi и др., 2015 г.).

Средние значения индексов на всех станциях в пределах пашенных земель конкретного сельскохозяйственного региона переписи (CAR) использовались для представления среднего агроклимата этого CAR. Если в CAR отсутствовали входные климатические данные, использовались станции из соседних CAR.

Чтобы сформировать управляемый массив предикторов потенциальной урожайности, AAFC объединило ежедневные агроклиматические индексы в месячные суммы и средние значения за период с мая по август. Их стандартные отклонения (Std) за месяц также были рассчитаны и включены в методологию моделирования (Newlands и др., 2014 г.; Chipanshi и др., 2015 г.). Значение Std показывает, как меняется дневной индекс в течение одного месяца. Чем больше Std, тем выше изменчивость параметра в этом месяце.

Выбор данных

Для обильных культур оценки урожайности на уровне CAR из июльских и ноябрьских обследований фермерских хозяйств с 1987 г. по настоящее время использовались в качестве входных переменных для моделей, в то время как оценки урожайности из сентябрьского и ноябряского обследований фермерских хозяйств использовались для проверки точности результатов модели урожайности. Для менее обильных культур данные обследования были собраны на уровне провинций.

Модель была выбрана путем сначала изучения существующих моделей, а затем путем оценки моделей, доступных в SAS. Моделирование проводилось на наименьшем географическом уровне, для которого были доступны исторические данные съемки. Были смоделированы только пять основных сельскохозяйственных провинций (Квебек, Онтарио, Манитоба, Саскачеван и Альберта).

После сравнения результатов с другими моделями, включая ступенчатую ненадежную модель и надежную модель наименьшего угла, Статистическое управление Канады приняло решение принять модель SAS LASSO. См. Статистическое управление Канады (2015 г.)²⁴ для получения более подробной информации.

²⁴ Интегрированное моделирование урожайности с использованием дистанционного зондирования, агроклиматических данных и данных обследований. http://www23.statcan.gc.ca/imdb-bmdi/document/5225_D1_T9_V1-eng.htm

Проблемы с качеством

Ноябрьское обследование фермерских хозяйств считается наиболее точной оценкой урожайности за данный год из-за того, что данные собирались после того, как большая часть сбора урожая завершена, а размер выборки является самым большим из всех шести обследований. Результаты сентябрьского обследования ферм можно считать предварительной оценкой ноябрьских результатов. Таким образом, цель моделируемой урожайности состоит не в том, чтобы воспроизвести результаты сентябрьского обследования фермерских хозяйств, а в том, чтобы получить достаточно точную оценку урожайности до получения результатов ноябрьского обследования.

Относительная разница (представленная в процентах) между оценкой урожайности данного метода (т. е. сентябрьского обследования ферм или модели урожайности) и оценкой урожайности ноябрьского обследования ферм была мерой точности. Отрицательная относительная разница указывала на то, что данная оценка урожайности была меньше оценки ноябрьского обследования фермерских хозяйств, а положительная относительная разница указывала на то, что данная оценка урожайности была выше оценки ноябрьского обследования фермерских хозяйств.

$$\text{Относительная разница} = 100 * \frac{\text{Оценка урожайности} - \text{Оценка урожайности ноябрьского обследования ферм}}{\text{Оценка урожайности ноябрьского обследования ферм}}$$

Более значимые относительные различия наблюдались в модельных оценках для культур, по которым имеется ограниченный объем исторических данных. Оценки, полученные на основе моделей, которые были построены только с ограниченным числом точек данных, подвергались риску быть статистически ненадежными.

Статистическое управление Канады установило три критерия на основе наличия входных данных, а также показателей качества, которые должны соблюдаться для обеспечения статистической достоверности оценок и определения того, какие из смоделированных урожаев были приемлемого качества для публикации на провинциальном и национальном уровнях. Для каждого года оценки моделей урожайности по каждой культуре будут оцениваться, чтобы определить, достаточно ли их качество для публикации.

Полученные результаты

Оценки, полученные с помощью модели SAS LASSO, были сопоставимы с оценками, полученными в ходе сентябрьского исследования, с точки зрения относительной разницы с оценками ноябрьского обследования для семи основных культур и многих из 12 дополнительных культур, опубликованных в сентябре. В редких случаях и модель, и сентябрьское обследование давали резкие относительные отличия от оценок ноябрьского обследования, но не обязательно для одних и тех же культур/годов. Эти крайние относительные различия, как правило, были больше для модели, чем для сентябрьского обследования.

Модель дала оценки урожайности с той же точностью, что и оценки, основанные на опросе. Смоделированные оценки удалось опубликовать раньше, что привело к экономии средств Статистического управления Канады и уменьшению нагрузки на респондентов. С 2016 года обследование фермерских хозяйств не проводится, а публикуются только оценки, основанные на модели.

5. Руководство для НСУ, рассматривающих возможность использования НЗ для официальной статистики

Данные НЗ и большие данные в более широком смысле представляют значительные возможности для дополнения и улучшения официальной статистики. Однако существуют также проблемы с использованием этих источников данных для статистических целей. Как указано в настоящем отчете, для НСУ важно определить алгоритмический подход и статистические приложения, которые больше всего подходят для использования данных НЗ, исходя из явных организационных преимуществ, осуществимости методов и вероятной экономии средств. Также стоит учитывать, что диапазон статистических приложений, которые извлекают выгоду из данных НЗ, скорее всего, будет со временем увеличиваться по мере увеличения качества, охвата, частоты, разрешения и пропускной способности каналов спутниковых изображений. Учитывая это, НСУ важно критически оценить, подходит ли использование больших данных в целом и данных НЗ в частности для их статистических целей.

В этой главе приводятся рекомендации, которые помогут НСУ выполнить такую оценку, однако она не предназначена для использования в качестве универсального руководства.

5.1 Потенциальное применение данных НЗ для официальной статистики

Данные НЗ и другие источники больших данных могут применяться для улучшения или замены некоторых процессов НСУ, которые производят официальную статистику. Tam & Clark (2015 г.) предлагают следующие приложения, проводя параллели с хорошо зарекомендовавшим себя использованием административных данных для официальной статистики:

- формирование основы выборки или регистра – определение единиц обследования и/или предоставление вспомогательной информации, такой как переменные стратификации;
- частичная подмена данных для подгруппы совокупности – уменьшение размера выборки;
- частичная замена некоторых необходимых элементов данных – сокращение длины инструмента обследования или обогащение набора данных без необходимости статистической увязки;
- вменение отсутствующих элементов данных – замена той же или аналогичной единицы;
- редактирование – помочь в обнаружении и обработке аномалий в данных обследования;
- связывание с другими данными – создание более полных наборов данных и/или продольных перспектив;
- сопоставление данных – обеспечение достоверности и непротиворечивости данных обследования; и
- генерирование новых аналитических идей – улучшение измерения и описания экономических, социальных и экологических явлений. (стр. 442).

Это исчерпывающий список. По мере того, как НСУ набираются опыта в использовании данных НЗ для официальной статистики, может появиться больше приложений.

5.2 Анализ затрат и результатов

Проведение анализа затрат и выгод полезно для принятия решения о том, следует ли использовать данные НЗ для статистических приложений.

Источники больших данных, включая данные НЗ, следует использовать только в том случае, если существует острая потребность бизнеса и, если это улучшит статистические результаты на основе объективных критериев затрат и выгод.

Согласно Tam (2015 г.), большие данные следует использовать только в том случае, если возможно:

- улучшить предложение продуктов статистических управлений, например, более частый выпуск официальной статистики, более подробная статистика, больше статистики по небольшим группам населения или районам или заполнение важных пробелов в данных;
- повысить экономическую эффективность производства официальной статистики – выгода для бизнеса (стр. 4).

Согласно Tam & Clark (2015 г.):

Затраты и выгоды от использования нового источника данных необходимо оценивать с точки зрения таких факторов, как снижение нагрузки на поставщика, устойчивость нового источника, а также точность, актуальность, непротиворечивость, интерпретируемость и своевременность выходных данных, предусмотренных в Системе оценки качества (стр. 8).

Приводится пример проведения этих оценок с использованием сельскохозяйственного пилотного проекта ABS (Австралийского бюро статистики), описанного в главе 4.

Полная замена данных, полученных на основе обследований, данными спутникового зондирования для производства сельскохозяйственных статистических данных, таких как растительный покров и урожайность, можно оценить следующим образом:

- **Затраты:** каковы вероятные затраты на получение, очистку и подготовку данных спутникового зондирования в форме, подходящей для дальнейшей официальной обработки, учитывая, что вычислительные потребности для получения, передачи, обработки, интеграции и анализа больших наборов данных изображений в настоящее время неизвестны, но могут со временем снизиться? Каковы затраты на разработку статистической методологии для преобразования данных спутникового зондирования в урожайность сельскохозяйственных культур; и на разработку статистической системы для обработки и распространения данных спутникового зондирования? Какова стоимость получения информации наземных ориентиров (например, данные наземных ориентиров ABS, охватывающие ~1 тыс. местоположений? Исследование LUCAS, охватывающее ~ 270 тыс. местоположений по всей Европе)? Каковы эквивалентные затраты на прямой сбор данных и как они соотносятся друг с другом?
- **Снижение нагрузки на провайдеров:** насколько уменьшится нагрузка на провайдеров, если прямой сбор данных будет заменен данными спутникового зондирования? Насколько важно такое сокращение, исходя из существующего опыта поставщиков и преобладающей политики правительства по сокращению регулятивной «бюрократизации»? Какова текущая степень сотрудничества фермеров и насколько вероятно, что эта степень изменится в лучшую или худшую сторону в будущем? Какие данные получают иначе, например, данные от фермеров, полученные в электронном виде, такие как урожайность?
- **Устойчивость статистических продуктов:** доступен ли источник данных официальным статистикам для регулярного производства официальной статистики? Какова вероятность того, что источник данных сохранится в будущем, а также сохранит текущий уровень качества?
- **Точность, релевантность, непротиворечивость, интерпретируемость и своевременность:** как новый источник данных соотносится с текущим источником в соответствии с критериями, изложенными в структурах качества данных (см. Главу 2)? И хотя оптическое спутниковое зондирование обеспечивает точные измерения «отражательной способности» – измеряя количества света, отраженного от Земли обратно до спутниковых датчиков, – существует еще ряд других вопросов.
- **Наличие знаний для обработки данных НЗ, а также специалистов по данным?**

Являются ли эти проблемы/неточности большими или меньшими, по сравнению с теми проблемами, которые связаны с отсутствием данных из прямых источников? Кроме того, преобразование коэффициента отражения в статистику растениеводства требует научного или статистического моделирования, которое обычно не используется НСУ и может вызвать проблемы с интерпретацией. С другой стороны, можно количественно оценить статистическую точность

прогнозных оценок, сделанных на основе этих моделей, равно как и оценить проблемы, связанные с качеством данных при получении этих оценок.

Поскольку данные спутникового зондирования доступны в среднем с любым заданным интервалом (от 15 минут в день до месяца), они имеют явное преимущество перед ежегодным или субгодовым прямым сбором данных с точки зрения частоты доступности статистических данных об урожайности (Tam & Clark, 2015 г., стр. 8).

В этом разделе описаны вопросы, которые были поставлены ABS и на которые были даны ответы для реализации своего пилотного проекта с использованием данных о НЗ для идентификации сельскохозяйственных культур, а также пример того, как другие НСУ могут подойти к решению о том, следует ли включать данные НЗ в свои статистические материалы.

5.3 Достоверность статистического вывода

Если определено, что использование данных НЗ удовлетворит потребности бизнеса или создаст эффективность, следует также рассмотреть, будет ли использование данных НЗ статистически достоверным. Согласно Tam & Clark (2015 г.), «предположение о том, что большие наборы данных каким-то образом ближе к «истине», не принимается статистиками, поскольку объективная «истина» очень сильно зависит от того, насколько репрезентативным является конкретный источник больших данных в части генеральной совокупности и характера статистических выводов, сделанных из таких данных» (стр. 437).

Недостаточный охват может быть проблемой для источников данных НЗ, поскольку, хотя наборы данных часто бывают большими, они не обязательно охватывают целевую группу населения, в отношении которой НСУ хочет сделать выводы. Например, если бы набор данных содержал данные НЗ по урожайности для Тасмании, это было бы недопустимо для выводов обо всей Австралии. Согласно Tam & Clark (2015 г.), в некоторых случаях «может потребоваться дополнить источники больших данных данными опросов, чтобы получить охват непредставленных сегментов населения. В других случаях может быть полезно опубликовать статистику, описывающую подгруппы» (стр. 442).

5.4 Выбор подходящего набора данных НЗ

При определении того, подходят ли данные НЗ для конкретных статистических целей, ключевой вопрос заключается в том, могут ли требуемые данные или информационные продукты вообще быть получены из НЗ. Дерево решений на Рисунке 33 поможет определить, уместно ли и возможно ли использовать продукты данных НЗ.

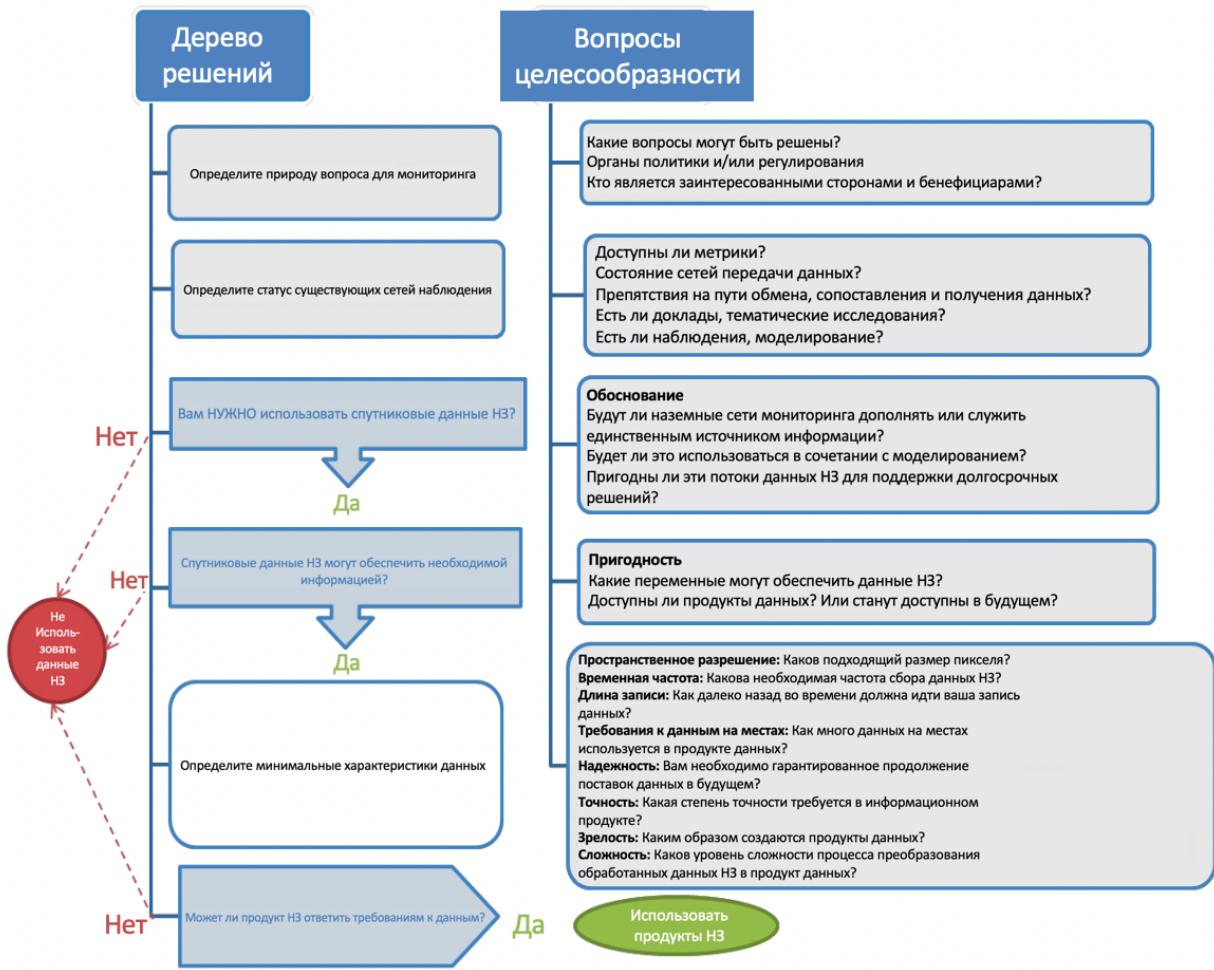


Рисунок 33: Схема принятия решений CSIRO по использованию данных НЗ для НСУ.

Тематические исследования внедрения данных НЗ для статистических целей были бы полезны для дальнейшей помощи в принятии решений для экономического обоснования относительно методологической осуществимости и затрат на получение, подготовку и предварительную обработку этих данных. Такие тематические исследования могут затем помочь в принятии решений для будущих приложений.

5.4.1 Минимальные требования к данным НЗ

Чтобы определить минимальные требования к данным НЗ, необходимо задать следующие вопросы и ответить на них, прежде чем принимать решение об использовании данных НЗ:

Обоснование	Нужно ли использовать НЗ? Есть ли лучший альтернативный источник?
Пригодность	Может ли НЗ предоставить необходимые информационные продукты?
Пространственное разрешение	Каков подходящий размер пикселя?
Временная частота	Какова необходимая частота сбора данных НЗ?
Длина записи	Как далеко назад во времени должна идти ваша запись данных?

Надежность	Вам необходимо гарантированное продолжение поставок данных в будущем?
Точность	Какая степень точности требуется в информационном продукте?
Зрелость	Вы хотите использовать только хорошо задокументированные и часто используемые информационные продукты?
Сложность	Какие возможности управления, обработки и анализа данных доступны?

Рисунок 34: Требования к использованию НЗ.

Помимо этих традиционных требований данным НЗ, необходимо учитывать и другие критерии.

5.4.2 Источники данных: доступ и владение

После того, как НСУ определило, что данные НЗ соответствуют их требованиям к выходным статистическим данным, необходимо определить источники и получить к ним доступ. В главе 2 описаны различные доступные источники, в том числе некоторые свободно доступные и частные ресурсы.

Основной риск для НСУ, включающих данные НЗ в свой процесс статистического производства, заключается в том, что они «имеют мало контроля или полномочий в отношении больших данных, находящихся в распоряжении частных организаций, и нет гарантии того, что компания и ее данные будут существовать в будущем. Если доступ будет запрещен в будущем или производство данных будет прекращено, то существует значительный риск для непрерывности поступления данных и наборов данных временных рядов, особенно если существующие системы были заменены новым решением для больших данных». (Kitchin, 2015 г.). В литературе обычно считается, что в настоящее время в целом отсутствует законодательная база и структура для НСУ в отношении владения данными и доступа к большим данным и данным НЗ.

Это соображение для НСУ, которые впервые начинают работать с данными НЗ на постоянной основе. Следует подумать о том, как повлияет на статистический процесс и результаты тот факт, если доступ к источнику данных НЗ станет невозможным или изменится содержание источника данных. Одним из примеров является внезапная потеря связи с эксплуатируемым более 10 лет спутником ENVISAT с 9 приборами на борту. Другим примером являются два датчика MODIS (Terra и Aqua), которые были задуманы как исследовательские инструменты, но оказались настолько успешными (и пережили свой ожидаемый срок службы в два-три раза), что обеспечили весьма объемный материал для работы. С появлением долгосрочных оперативных спутниковых миссий Copernicus Sentinel, финансируемых Европейским Сообществом и управляемых ESA и ЕВМЕТСАТ, теперь появилась возможность инвестировать в методы использования этих сложных данных о Земле как минимум в течение следующих 15 лет.

5.5 Оценка качества

Качество данных зависит не только от их собственных характеристик, но и от того, соответствуют ли они релевантным целям и бизнес-требованиям (Cai & Zhu, 2015 г.). Таким образом, качество является многомерным понятием, которое включает не только точность статистических данных, но и другие аспекты, такие как: актуальность, своевременность, точность, согласованность, интерпретируемость и доступность. Поэтому оценка качества должна включать эти аспекты.

Как правило, продукты, полученные с помощью дистанционного зондирования, должны иметь показатель оценки точности, который указывает, в какой степени данные правильно описывают явление, для измерения которого они предназначены. Для оценки точности может потребоваться сравнение результатов с другими независимыми (но также подверженными ошибкам) источниками информации, такими как: наземные справочные данные, административные регистры и т. д.

Согласно рекомендациям по качеству данных GEOSS (2013 г.), поставщики данных должны «учитывать многочисленные аспекты качества. Цель состоит не в оценке или ранжировании ресурсов данных, а в описании характеристик, которые необходимо знать, чтобы пользователь мог решить, следует ли их использовать» (стр. 1).

Эти характеристики содержат (см. Таблицу 1 в Главе 2):

- Охват, включающий как пространственные, так и временные измерения;
- Последовательность, в том числе долгосрочная последовательность;
- Неопределенности оценены и задокументированы, включая как пространственные, так и временные измерения;
- Атрибуция источников ошибок (частота и причина отсутствия значений);
- Информация о валидации, т. е. как данные были оценены на предмет неопределенностей по сравнению с альтернативными измерениями;
- Задержка со времени проведения наблюдения;
- Разрешение, включая как пространственное, так и временное измерение;
- Юзабилити в том смысле, что форма и содержание удобны для использования в конкретных целях;
- Простота, так что данные и базовая модель данных не являются излишне сложными для понимания и манипуляций (GEOSS, 2013 г., стр. 2);

5.6 Распространение результатов

Стратегия распространения, реализуемая НСУ, должна способствовать доступу, интерпретации и повторному использованию данных. Эти факторы помогают усилить использование информации.

Некоторые аспекты, которые можно было бы учитывать при распространении результатов на основе данных НЗ:

- Оценки, основанные на статистических моделях и других подходах, изложенных в Главе 3 и дополнительных материалах, должны давать связанную меру статистической точности оценки с использованием надежных методов, которые четко задокументированы.
- Методология, лежащая в основе статистической модели, исходные допущения и их возможные последствия, качество данных НЗ и процессы, предпринятые для подготовки данных, как показано в главе 2, должны быть полностью задокументированы в руководстве пользователя.
- Визуализация данных: Обычно продукты, связанные с обработкой данных НЗ, имеют пространственный компонент (карты, изображения, географические слои). Использование географических средств просмотра позволяет запрашивать эту информацию.
- Интерпретируемость: с целью повышения полезности информации рекомендуется предоставлять метаданные, которые описывают основные характеристики используемого изображения (дата захвата, пространственное разрешение, датчик), процедуры, меры точности и другие аспекты, которые считаются важными.
- Своевременность: чтобы сократить «время вывода на рынок» статистических продуктов, задержка между отчетным периодом (к которому относятся данные) и датой, когда данные становятся доступными, должна быть минимальной.
- Непатентованный формат: если продукты доступны для скачивания, рекомендуется публиковать их в непатентованном формате.
- Правовые аспекты: Следует гарантировать, что публикуемая информация соответствует правовым аспектам, установленным в стране.

5.7 Другие соображения и вопросы

В дополнение к проблемам, изложенным ранее в этой главе, следует также учитывать следующее, если НСУ решит работать с данными НЗ:

- Технологические усовершенствования и достижения, такие как более сложные датчики, со временем приведут к появлению новых источников данных НЗ. НСУ должны регулярно пересматривать и обновлять свои источники данных для получения статистических результатов по мере появления более качественных источников.
- Выявление конкретного проекта или группы проектов, которые могут быть решены в разумные сроки. При таком подходе НСУ стремится к управляемым результатам, которые, скорее всего, получат поддержку со стороны высшего руководства.
- Оценка ресурсов, которые могут быть развернуты. Определяющими факторами являются такие вопросы, как бюджет, аппаратное обеспечение, устройства хранения данных, программное обеспечение и информационная политика.
- Объективная оценка знаний и навыков, имеющихся в организации, является ключевым фактором. По данным Рабочей группы по большим данным правительства Австралии (2015 г.): «Существенным препятствием для реализации ценности больших данных будет нехватка талантов, особенно людей с глубокими знаниями в области статистики и машинного обучения, а также менеджеров и аналитиков, которые знают, как управлять компаниями, используя информацию из больших данных». (стр. 12). Решения по обучению и развитию сотрудников повышают устойчивость инициатив и их успех. Как упоминалось в главе 2, существует гораздо больше подходов к использованию данных НЗ, чем только статистический подход, и в долгосрочной перспективе эти подходы, более основанные на физике (включая асимиляцию модельных данных), могут быть более поддающимися автоматизации и, таким образом, способствовать снижению затрат. Тогда возникает вопрос, как НСУ взаимодействуют с теми экспертными организациями, которые обладают необходимыми возможностями.
- Когда статистические данные и информация имеют географическую привязку или пространственную привязку, требуются новые навыки или их дальнейшее развитие. Навыки пространственного мышления и анализа должны быть улучшены или приобретены, чтобы отвечать на такие вопросы, как размер и распределение выборки, статистический и географический охват, географические уровни, сбор и распространение данных, а также конфиденциальность данных, среди прочего.
- Активное участие в форумах, таких как Глобальная рабочая группа ООН по большим данным, облегчает переход к использованию источников больших данных и данных НЗ. Доступ к практическому опыту, методам, тенденциям и постоянные обсуждения являются очень важными инструментами для поддержания «живых» проектов, основанных на данных НЗ. Рекомендуется партнерство или сотрудничество с двумя международными организациями, которые используют и координируют данные НЗ, такими как GEO и CEOS.

5.8 Дополнительные рекомендации для специалистов, работающих со своими первыми данными наблюдений Земли.

Наращивание потенциала посредством сотрудничества с международными и национальными агентствами.

Начать работать с данными НЗ в первый раз может быть сложно, поэтому НСУ должны учиться на знаниях и опыте других агентств. По мере того, как они приобретают опыт работы с данными НЗ, НСУ должны делиться этими знаниями с другими организациями в своем окружении, со временем наращивая потенциал заинтересованных сторон во всем мире. Партнерские отношения между поставщиками данных, НСУ, частным сектором и академическими кругами важны для дальнейшего использования данных НЗ для целей официальной статистики. В тех случаях, когда НСУ не имеет опыта работы с такими областями, как геодезические системы отсчета, картографические проекции, растровые и векторные форматы и т. д., ситуация с данными может быть сложной. Первым шагом

для выхода из ситуации является укрепление связей с национальным картографическим агентством страны.

Интенсивные курсы

Проверенным и успешным методом обучения экспертов НСУ может быть создание интенсивных курсов продолжительностью от одной до трех недель с целью повышения уровня необходимых знаний в области НЗ до необходимого уровня для успешного использования этих данных в статистических целях.

Использование информации из открытых источников в Интернете

В настоящее время много полезной информации доступно на веб-сайтах, в блогах, видео, форумах пользователей и в бесплатных курсах. Такой обмен информацией может дать большое преимущество, и можно сказать, что с помощью этих инструментов можно решить большинство практических вопросов.

Определение навыков, необходимых организации

По мере накопления опыта работы с данными НЗ НСУ могут выявлять внутренние пробелы в навыках, когда требуются люди с более высокой или другой квалификацией. Например, программы наращивания организационного потенциала могут включать, среди прочего, обучение в области науки о данных, инфраструктуры пространственных данных, дистанционного зондирования, анализа пространственных данных и информатики. Если у национальной статистической службы нет необходимых внутренних навыков, она также может рассмотреть вопрос о передаче некоторых задач на аутсорсинг другим организациям, обладающим соответствующим опытом, или сформировать междисциплинарные группы в агентствах.

Взаимодействие с общественностью для улучшения восприятия конфиденциальности

При внедрении использования больших данных, таких как данные НЗ, важно взаимодействовать с общественностью, чтобы гарантировать, что ее доверие к НСУ и результатам ее статистической деятельности не пострадает. В сообществе могут возникнуть опасения, что при использовании и объединении источников больших данных для официальной статистики их конфиденциальность будет нарушена. Для НСУ «крайне важно, чтобы эти проблемы решались с помощью таких методов, как обеспечение прозрачности в отношении того, как и какие используются источники больших данных» (Strujs, Braaksma & Daas, 2014 г.). Данные НЗ являются одной из наиболее приемлемых форм больших данных, поскольку они обычно используются Google Earth и прогнозами погоды (применяя ассимиляцию моделей, основанных на физике, с локальными данными и данными наблюдения Земли с метеорологических спутников), включая дождевые радары.

Учет этических принципов и действующего законодательства о конфиденциальности

Поскольку в настоящее время законодательство не было обновлено для всестороннего рассмотрения дел, связанных с большими данными и проблемами конфиденциальности, юридические обязательства НСУ не всегда ясны. Предотвращение «раскрытия личности людей является императивом, но это трудно гарантировать при работе с большими данными» (Strujs, Braaksma and Daas, 2014 г.). НСУ имеют свои собственные политики и полномочия в отношении конфиденциальности и защиты информации, которую они собирают, и проявляют разумную осторожность для защиты личности людей, в этом плане учет этических стандартов может служить руководством. По мере обновления законодательства и руководящих принципов НСУ должны обновлять свою политику, чтобы отразить эти изменения. Этот аспект относится главным образом к использованию спутниковых данных очень высокого (субметрового) разрешения.

Возможность использования готовых к анализу данных (ARD)

Обработка, необходимая для преобразования необработанных данных НЗ для подготовки к анализу, может занимать много времени и быть дорогостоящей. Для НСУ, не знакомых с такой практикой и не имеющих необходимых навыков, затраты на этот процесс могут быть непомерно высокими. Если это так, НСУ должны оценить, будет ли приобретение ARD более эффективным с точки зрения затрат, либо путем первоначального поиска ARD, либо путем аутсорсинга обработки необработанных данных. Дополнительную информацию об ARD для статистических целей см. в разделе 2.8 настоящего отчета.

Распространение анализа результатов и алгоритмов

Распространяя результаты анализа данных НЗ, НСУ могут помочь другим организациям развивать свои навыки работы с данными НЗ, предоставляя примеры, получая отзывы и советы от коллег и принося пользу обществу посредством статистических материалов, получаемых при меньшей нагрузке на поставщиков.

По мере проведения дальнейших исследований в области использования наблюдений Земли в статистических целях, если заинтересованные стороны продолжат делиться своими знаниями и опытом с другими коллегами через глобальные каналы, такие как Глобальная рабочая группа ООН по большим данным и Комитет экспертов ООН по управлению глобальной геопространственной информацией (UN GGIM), работа в этой области будет продолжать развиваться и приносить пользу мировому сообществу благодаря таким примерам, как измерение численности населения и мониторинг ЦУР. Воспроизводимость результатов за счет совместного использования рабочего процесса/сценариев значительно улучшает их.

6. Заключение

Использование данных НЗ для официальной статистики представляет как возможности, так и проблемы для НСУ. В этом отчете содержится информация, которая поможет специалистам-практикам в процессе критической оценки своих бизнес-потребностей в данных НЗ, рассмотрении источников, методологических подходов, примеров тематических исследований и руководящих принципов, которыми следует руководствоваться при принятии решений об использовании данных НЗ в статистических целях.

Использование данных НЗ должно осуществляться, когда есть явная выгода для бизнеса, и это необходимо оценивать в каждом конкретном случае. Учитывая эту изменчивость в отношении того, когда и как целесообразно использовать данные НЗ, в будущем обмен информацией, опытом и извлеченными уроками между правительствами, частным бизнесом и исследовательским сообществом будет иметь важное значение.

С этой целью инициативы Глобальной рабочей группы СОООН по большим данным и ее целевых групп, UN GGIM, CEOS и GEO будут и впредь обеспечивать ценную международную сеть профессионалов, обменивающихся знаниями и развивающих потенциал использования источников больших данных, таких как данные НЗ для статистического и политического применения. Наиболее заметным использованием станет измерение, мониторинг и отчетность о ходе достижения Повестки дня в области устойчивого развития на период до 2030 года.

По мере того, как использование данных НЗ становится все более распространенным, а методы использования этих источников совершенствуются, могут стать очевидными дальнейшие сферы применения и расширяться те области, которые в настоящее время не предвидятся. На протяжении всего этого процесса расширения, открытая коммуникация и обмен знаниями будут оставаться важными, особенно в интересах практиков, впервые внедряющих источники данных НЗ в работу своей организации.

Приложение. Обзор статистических и машинных подходов к обучению

Хотя статистики и знакомы со стандартными статистическими моделями, они могут быть не так хорошо знакомы с методами машинного обучения. Цель данного Приложения – описать ряд моделей машинного обучения, которые использовались при анализе данных НЗ, указать примеры их использования в этом контексте и предоставить ссылки для дальнейшего изучения. Методы классифицируются в соответствии с четырьмя типами аналитических целей, подробно описанных в главе 3, а именно: классификацией, кластеризацией, регрессией и уменьшением размерности. Они также классифицируются в зависимости от того, на что сделан акцент: на анализ за какой-то один период времени (статистические данные) или за целый период времени (динамические данные).

A.1 Методы анализа статистических данных Методы классификации

Логистическая и полиномиальная регрессия являются частью семейства обобщенных линейных моделей (glm). Другими примерами обобщенных линейных моделей, которые можно использовать для классификации, являются регрессия Пуассона (например, для данных подсчета) и отрицательная биномиальная регрессия (например, для более редких пространственных результатов). Логистическая регрессия используется, когда выходная переменная (отклик) является категориальной с двумя уровнями (растительность/не растительность, высокая/низкая, присутствует/отсутствует). Полиномиальная логистическая регрессия используется, когда переменная отклика имеет уже более двух уровней (деревья/трава/голая почва/урожай/вода, высокий/средний/низкий).

Если имеется большое количество входных переменных или некоторые переменные потенциально избыточны, можно использовать методы регуляризации, чтобы уменьшить или устраниć влияние переменных, которые не вносят существенного вклада в анализ. Общие методы включают регуляризацию ridge- и лассо-, а также более общую регуляризацию эластичной сети. Другие методы выбора переменных включают знакомые методы прямой, обратной и поэтапной регрессии, известные в литературе по машинному обучению как построение прямых признаков и устранение обратных признаков (Silipo, 2015 г.). Однако эти подходы требуют больших затрат времени и вычислительных ресурсов и практически применимы только к набору данных с уже относительно небольшим количеством входных переменных.

Пример использования логистической регрессии с данными НЗ привел Bavaghar (2015 г.). Автор применяет метод для оценки местоположения и степени обезлесения на основе таких переменных, как уклон и расстояние до дорог и жилых районов, и подчеркивает, в частности, возможность количественной оценки неопределенности в прогнозах как сильную сторону такого подхода. Был получен общий уровень правильной классификации равный 75% с оценкой обезлесения 12% всего исследуемого участка за 27 лет с 1967 по 1994 годы.

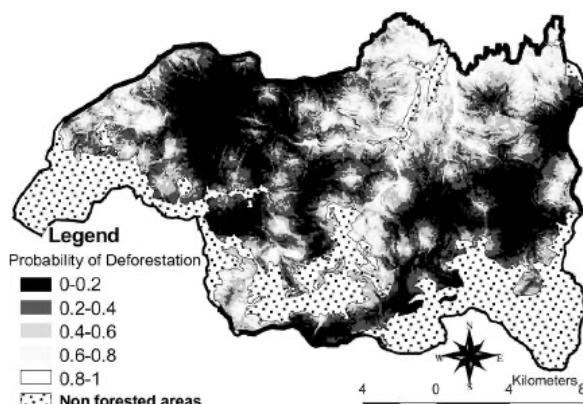


Рисунок 35: Вероятность обезлесения, полученная с помощью логистического регрессионного анализа данных НЗ. Bavaghar (2015 г.)

Классификация методом максимального правдоподобия по Гауссу является одним из наиболее широко используемых методов контролируемой классификации данных спутниковых изображений (Richards, 2013 г.). Цель состоит в том, чтобы классифицировать наблюдения по группам, где считается, что наблюдения в каждой группе имеют гауссово распределение. Также доступны более гибкие альтернативы, основанные на полупараметрической оценке ядра. Также утверждается, что метод максимального правдоподобия эквивалентен линейному или квадратичному дискриминантному анализу (Hoglan и др., 2013 г.).

Нейронные сети являются популярными методами классификации. Их также можно использовать для регрессии (см. ниже). При графическом представлении нейронная сеть состоит из нескольких слоев. Первый слой – это входной слой, представляющий каждую из переменных-предикторов. За этим следует один или несколько слоев скрытых узлов, каждый из которых представляет функцию активации, действующую на взвешенные входные данные выходных данных предыдущих слоев. Наконец, выходной слой может быть одним слоем в случае регрессии или в случае классификации будет состоять из узла для каждой возможной выходной категории. Из-за плотной связности узлов в нейронной сети окончательная функция может представлять сложные отношения между входными переменными, а также между входными и выходными переменными.

Нейронные сети обычно «обучаются» или оцениваются количественно с помощью алгоритма обратного распространения, который похож на градиентный спуск, и итеративно корректирует веса графика после просмотра каждой новой точки данных. Это обновление весов приводит нейронную сеть к некоторым локальным минимумам в пространстве параметров в зависимости от точности обучения. Для эффективного обучения нейронных сетей обычно требуется относительно большой набор обучающих данных.

Глубокое обучение относится к разработке новых алгоритмов для преодоления проблем, связанных с обучением нейронных сетей, которые имеют много скрытых слоев. В частности, проблема «исчезающего градиента», которая относится к ограничению традиционного обратного распространения при обновлении весов более ранних слоев в глубоких сетях.

Нейронные сети использовались для анализа данных НЗ почти два десятилетия; см. исследования Chen и др. (1995 г.), Paola & Schowengerdt (1995 г.), Atkinson & Tatnall (1997 г.), Paola & Schowengerdt (1997 г.) & Foody и соавторы (1995 г., 1997 г., 2001 г.), Kavzoglu & Mather (2003 г.), Mertens и др. (2004 г.), Verbeke и др. (2004 г.), Erbek и др. (2004 г.), Kussul и др. (2006 г.) & Kavzoglu & Reis (2008 г.). Более свежие примеры классификации с использованием нейронных сетей в контексте данных НЗ приведены у Srivastava и др. (2012 г.), Shao & Lunetta (2012 г.) и Wang и др. (2015 г.).

López-Granados и др. использовали методы нейронной сети для выявления различий в отражательной способности между пшеницей и злаковыми сорняками (2008 г.). Авторы использовали предварительное вычисление наиболее релевантных главных компонентов, чтобы значительно повысить точность классификации до 90–100 %.

Деревья классификации – это метод контролируемой классификации, который представляет принадлежность к классу как «листья», а входные переменные – как «узлы». Из этих узлов формируются ветви на основе разбиения значений входных переменных для наилучшей группировки данных.

Дерево классификации является членом семейства методов дерева решений, которые включают в себя деревья регрессии, бустинг деревьев, бэггинг деревьев и т. д. Деревья классификации аналогичны деревьям регрессии (см. ниже), разница заключается в распределении ответов (категориальное или непрерывное). Деревья классификации и регрессии (CART) строятся с использованием подхода рекурсивного разбиения, при котором на каждом этапе выбираются переменные, которые лучше всего разбивают данные на однородные классы. Критерием выбора переменных и принятия решения о разделении обычно является мера однородности подмножества, такая как индекс Джини или энтропия, или, в случае деревьев регрессии, уменьшение дисперсии.

Основное преимущество деревьев классификации в частности и деревьев решений в целом заключается в том, что они просты и понятны. Благодаря своей вычислительной простоте они могут быть применены к большим объемам данных. Однако деревья решений могут страдать от чрезмерной подгонки и часто требуют обрезки, что, по сути, является проблемой выбора модели. Более того, нет никакой гарантии, что будет найдено глобально оптимальное дерево. Популярные расширения деревьев решений, предназначенные для решения этих проблем предвзятости и дисперсии, включают бустинг (построение последовательности деревьев, каждое из которых построено на неправильно классифицированных наблюдениях предыдущего дерева) и бэггинг (построение набора деревьев на основе случайных наборов данных). Случайный лес (RF) – это набор «мелких» деревьев, построенных из множества случайных выборок.

Деревья классификации использовались для анализа данных НЗ более 15 лет. Например, Lawrence & Wright (2001 г.) использовали этот метод для классификации земного покрова в экосистеме Йеллоустона в США на основе снимков Landsat TM. Входные переменные дистанционного зондирования были извлечены из двух сцен TM, охватывающих регион, и включали каналы 1-7 для каждого набора данных, а также яркость, зелень и влажность. Высота, уклон и экспозиция были извлечены из цифровой карты высот (DEM), а 30-100 эталонных участков были закодированы по аэрофотоснимкам. «Классификация на основе правил» была получена на основе анализа дерева классификации; выдержка из этой схемы показана ниже.

```
1. Tasseled Cap brightness difference < 22.5
  1.1. Elevation < 1682 m
    1.1.1. June TM band 6 < 139.5
      1.1.1.1. June TM band 4 < 111, THEN Natural vegetation
      1.1.1.2. June TM band 4 > 111, THEN Natural vegetation
    1.1.2. June TM band 6 > 139.5
      1.1.2.1. June TM band 2 < 32.5, THEN Agriculture
      1.1.2.2. June TM band 2 > 32.5, THEN Agriculture
  1.2. Elevation > 1682 m
    1.2.1. June TM band 6 < 116.5, THEN Natural vegetation
    1.2.2. June TM band 6 > 116.5, THEN Natural vegetation
2. Tasseled Cap brightness difference > 22.5
  2.1. Elevation < 1754 m
    2.1.1. Tasseled Cap greenness difference < 1
      2.1.1.1. June TM band 4 < 116.5
        2.1.1.1.1. Tasseled Cap wetness difference < -5.5
        2.1.1.1.1.1. August TM band 5 < 138.5
```

Рисунок 36: Фрагмент структуры принятия решений на основе правил, построенной с использованием дерева классификации на основе данных НЗ. Lawrence & Wright (2001 г.).

Более свежие ссылки на деревья решений и дистанционное зондирование включают Sharma и др. (2013 г.), Al-Obeidat и др. (2015 г.), Chasmer и др. (2014 г.), Otupei & Blaschke (2010 г.) и Peña-Barragán и др. (2011 г.). Как обсуждалось в тематическом исследовании в главе 4 настоящего отчета, Schmidt и др. (2016 г.) сравнили ряд методов классификации урожая/отсутствия урожая в Австралии с использованием изображений Landsat и обнаружили, что метод машинного обучения «случайные леса» обеспечивает наилучшую точность и надежность.

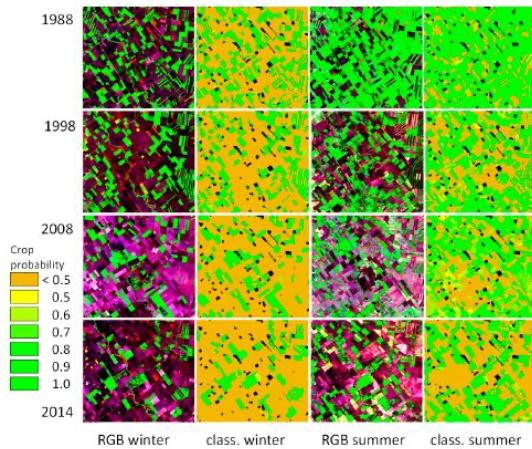


Рисунок 37: Классификация урожая/отсутствия урожая, полученная на основе анализа случайного леса изображений Landsat. Schmidt и др. (2016 г.).

Метод опорных векторов (SVM) – это класс непараметрических методов контролируемой классификации. В своей простейшей форме SVM с двумя классами представляют собой линейные двоичные классификаторы. Термин «опорный вектор» относится к точкам, лежащим на границе разделения между группами или классами данных.

Линейная разделимость часто недостижима (до желаемого уровня точности ошибочной классификации) в случае практического применения данных, поэтому были предложены различные методы для преодоления этого ограничения. Одним из наиболее распространенных является использование ядер для сопоставления данных с пространством более высокого измерения в надежде, что в этом пространстве данные станут разделимыми (или, по крайней мере, разделимыми до разумного уровня ошибочной классификации). Есть много доступных потенциальных ядерных функций, хотя наиболее распространеными в литературе по классификации сельскохозяйственных культур, по-видимому, являются радиальная базисная функция и полиномиальные ядра. Затем алгоритм SVM строит (линейную) гиперплоскость в этом многомерном пространстве для оптимального разделения групп, а потом отображает эту гиперплоскость обратно в исходное пространство. Результатом для изображения, например, является кривая, которая распределяет наблюдения по классам в зависимости от того, на какую сторону линии они лежат.

Отличным справочником по использованию SVM для дистанционного зондирования является публикация Mountrakis и др. (2011 г.). Некоторые другие документы, в которых исследуется использование SVM для анализа данных НЗ, включают Szuster (2011 г.) – классификация землепользования и растительного покрова в тропических прибрежных зонах; Mathur & Foody (2008 г.) – классификация сельскохозяйственных культур; Huang и др. (2002 г.) – классификация земного покрова; Melgani & Bruzzone (2004 г.) – данные дистанционного зондирования; и Shao & Lunetta (2012 г.) – классификация земель.

В некоторых из этих статей сравнивались различные подходы к SVM. Например, в статье Shao & Lunetta (2012 г.) предпочтение было отдано моделям CART из-за скорости и простоты использования, а также SVM и NN (нейронным сетям) – из-за точности классификации.

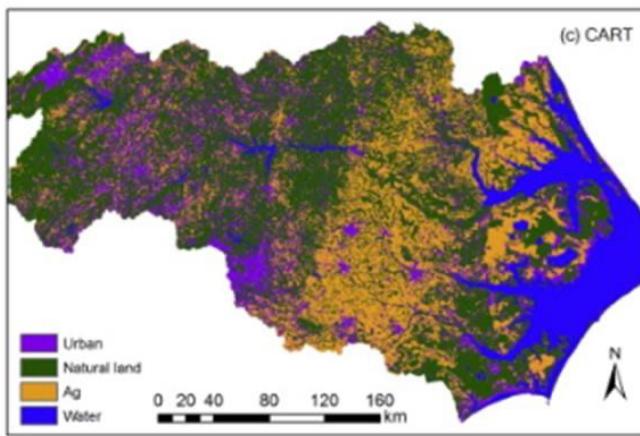


Рисунок 38: Карта классификации земель, созданная на основе анализа данных НЗ с помощью CART. Shao & Lunetta (2012 г.).

Методы ближайших соседей (k -nn) использовались более полувека (Fix & Hodges Jr, 1951 г.). Они являются хорошо известным и популярным методом непараметрической классификации из-за своей относительной простоты. В их простейшей форме объект классифицируется в соответствии с большинством голосов к его ближайших соседей. То есть, класс объекта присваивается ему наиболее распространенным классом среди k его ближайших соседей, где k обычно представляет собой небольшое положительное целое число.

Вклады соседей в класс объекта могут быть взвешены (например, обратной величиной их расстояния до объекта). Расширения метода, допускающие варианты «сомнения», известны как правила (k , 1): эти правила используют голос среди классов к ближайших соседей, но не выделяют класс большинства для объекта, если он не получает 1 или более голосов (иначе класс объявляется «сомнительным»).

Классификация с использованием правила k -nn может сильно зависеть от расстояния, используемого для определения ближайших соседей объекта, а также от выбора k . Кроме того, k -nn известен еще как «ленивый ученик»: он не учится на обучающих данных, а просто использует обучающий набор для классификации новых объектов. Из-за этого метод может сильно зависеть от локальных структур данных, и могут возникнуть большие затраты памяти для хранения адекватных обучающих данных, что делает его медленным. Однако редактирование данных может несколько улучшить эти проблемы, используя идею о том, что во многих случаях для очень хорошей аппроксимации границы решения k -nn необходимо лишь небольшое подмножество обучающих данных. Существует много алгоритмов редактирования данных: некоторые примеры см. в Ripley (1996 г.).

Некоторые ранние примеры работы в литературе по дистанционному зондированию с использованием подходов k -nn включают Hardin & Thomson (1992 г.), Hardin (1994 г.), Franco-Lopez и др. (2001 г.), Ohmann & Gregory (2002 г.), Collins и др. (2004 г.), Haapanen и др. (2004 г.), Zhu & Basir (2005 г.), Triepke и др. (2008 г.) и Blanzieri & Melgani (2008 г.).

Например, McRoberts и др. (2007 г.) использовали методологию k -nn для картирования и оценки лесоустройства. Данные НЗ включали изображения Landsat за 3 даты в 1999-2001 гг., соответствующие раннему, пиковому и позднему озеленению. Снимки использовались для получения 12 переменных-предикторов на основе спутниковых изображений, NDVI и преобразований ТС (яркость, зелень, влажность). Затем эти прогностические переменные использовались для оценки четырех атрибутов леса: доля площади леса (PFA), объем ($m^3/га$) (VOL), базальная площадь ($m^2/га$) (BA) и густота стволов (число деревьев/га) (D). Несколько пикселей были объединены для получения оценок (среднее значение, дисперсия) для интересующей области (АО). Авторы обнаружили, что некоторыми преимуществами подхода k -nn являются оценка по малой площади, многомерная оценка, а также точечная оценка и неопределенность.

Хотя методы ближайших соседей концептуально привлекательны и быстры в вычислительном отношении, они не всегда являются лучшей моделью для работы с данными НЗ. Например, ver Hoef & Temesgen (2013 г.) сравнили k-пп и пространственную линейную модель для использования в целях лесного хозяйства и обнаружили, что пространственная модель предпочтительнее. Однако ее точность и прогностические возможности можно улучшить, комбинируя ее с другими моделями. Например, Blanzieri & Melgani (2008 г.) предпочли комбинацию методов k-пп и SVM для классификации изображений дистанционного зондирования.

Подходы внутрипиксельной или субпиксельной классификации могут использоваться для решения проблемы так называемых «смешанных» пикселей: пикселей, которые отображают характеристики более чем одной группы. Смешанные пиксели в основном возникают в данных дистанционного зондирования с низким (например, MODIS) или средним (Landsat) разрешением. В контексте идентификации культур смешанные пиксели могут содержать, например, границы поля.

Двумя наиболее распространенными подходами в литературе по дистанционному зондированию в отношении проблемы классификации смешанных пикселей являются спектральный анализ совокупностей (SMA) и мягкие, или нечеткие, классификаторы. По обоим этим методам имеется огромное количество литературы. Нечеткие классификаторы и их варианты описаны в недавней книге Mather & Tso (2016 г.) наряду со многими другими методами классификации данных дистанционного зондирования. Подробности SMA и его сравнительные преимущества, а также недостатки для анализа данных дистанционного зондирования обсуждаются в книге Thenkabail (2015 г., стр. 229). Этот метод доступен во многих программных пакетах, например, в Wiki Landscape Toolbox:

http://wiki.landscapetoolbox.org/doku.php/remote_sensing_methods:spectral_mixture_analysis

Пример внутрипиксельной классификации приводится у Zhang и др. (2015 г.), которые предлагают стратифицированный временной спектральный анализ (STMSA) для оценки площади пахотных земель с использованием данных временного ряда MODIS.

Байесовские сети (BN) также чаще используются с данными НЗ для классификации. BN направлены на разработку набора отношений между входными переменными и выходной переменной (отклик), используя условную зависимость каждого атрибута от каждого другого атрибута. Результирующая модель может быть отображена графически с входными переменными в виде узлов и с зависимостями между переменными в виде направленных стрелок, соединяющих узлы. Каждый узел количественно определяется узлами, которые на него влияют, так что результирующая модель представляет собой последовательность условных вероятностей, ведущую к общей вероятности ответа. Сеть может информировать об относительной важности переменных, влияющих на реакцию, влияния на реакцию из-за изменений во входных переменных с учетом всех других переменных в модели и об оптимальных значениях входных переменных, ведущих к наилучшему ответу. BN используются в анализе данных НЗ уже более десяти лет. Например, Stassapoulou и др. (1998 г.) применили BN к данным ГИС и продемонстрировали дополнительное преимущество простого объединения нескольких источников данных в сложной структуре моделирования. Более свежий пример приводится у Mello и др. (2013 г.), которые использовали BN для обнаружения урожая сои с использованием данных дистанционного зондирования, подчеркнув возможность включения экспертной информации в модель для улучшения классификации.

Методы кластеризации

Метод k-средних – один из наиболее распространенных подходов к кластеризации, используемый в машинном обучении. Алгоритм предполагает, что данные берутся из K различных кластеров, и назначает каждую немаркованную точку ближайшему центру группы, которые пересчитываются до тех пор, пока не произойдет никаких изменений.

Метод k-средних можно использовать для различных целей. Например, он часто используется в качестве способа уменьшения размерности для анализа данных НЗ: выбирается значение k, которое велико, но намного меньше, чем исходное количество пикселей, а полученные кластеры затем

используются для дальнейшей классификации, регрессии или другого анализа. В качестве альтернативы кластеры, полученные в результате анализа k-средних, могут быть проверены, чтобы получить представление о важных факторах, которые различают группы, или для постфактум классификации. Например, Usman (2013 г.) использовал метод k-средних для классификации данных спутниковых изображений высокого разрешения по классам, которые затем были определены как сельскохозяйственные угодья, пустынные земли и застроенные территории.

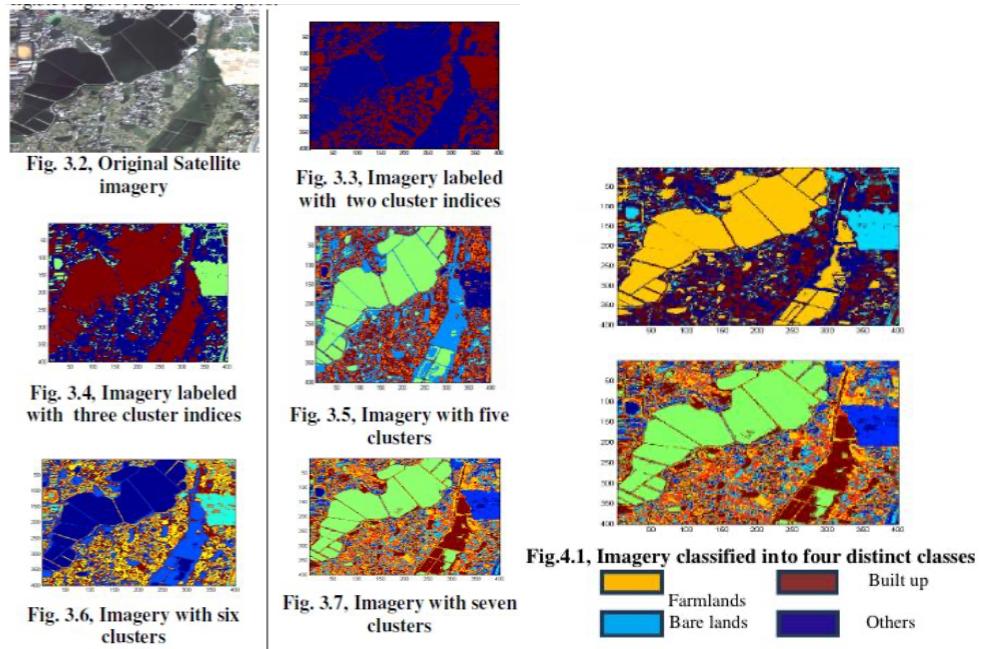


Рисунок 39: Пример кластеризации k-средних спутниковых данных с разными значениями k. Кластеры классифицируются по типам землепользования. Usman (2013 г.).

Агломерационная кластеризация и ее разновидности (например, иерархическая агломерационная кластеризация) также являются очень популярным подходом к кластеризации. Алгоритмы начинают с каждой точки как отдельного кластера и итеративно объединяют ближайшие кластеры, пока не будет достигнуто правило остановки.

Смешанные модели являются популярным методом кластеризации как в статистическом (эмпирическом) сообществе, так и в сообществах машинного обучения. Эти модели также известны как мягкие или нечеткие классификаторы.

Смешанные модели основаны на предположении, что наблюдаемые данные поступают из различных источников или групп. Предполагается, что каждая группа имеет определенное распределение; например, в смешанной модели Гаусса наблюдения в группе могут быть нормально распределены со средним значением и дисперсией, специфичными для группы. Затем смешанная модель представляет собой взвешенную сумму этих распределений, где веса соответствуют доле наблюдений в популяции, принадлежащих к этой группе; это также можно интерпретировать как вероятность того, что наблюдение принадлежит к этой конкретной группе. Таким образом, в «мягкой классификации» наблюдения распределяются более чем по одной группе с определенной вероятностью; это отличается от «жестких классификаторов», таких как kNN или k-средних (см. Раздел о методах машинного обучения ниже), в которых наблюдение относится только к одной группе.

В смешанной модели количество групп, веса и параметры распределений, которые описывают группы (например, средние значения и дисперсии компонентов гауссовой смешанной модели), могут быть известны или неизвестны. Затем анализ пытается оценить неизвестные компоненты модели и, таким образом, расшифровывает данные по составляющим группам. Затем они интерпретируются как кластеры.

Другие примеры использования смешанных моделей для анализа данных НЗ включают работу de Melo и др. (2003 г.) о контролируемой классификации мультиспектральных изображений дистанционного зондирования в районе реки Тапажос в Бразилии и Walsh (2008 г.), который объединяет вторичные оценки лесов, полученные на основе данных земного шара, и обследования домохозяйств, чтобы охарактеризовать причины и последствия лесовосстановления в районе Северной Эквадорской Амазонии.

Совсем недавно Tau и др. (2016 г.) использовали смешанную модель Гаусса для оценки и картирования городского земельного покрова с использованием изображений дистанционного зондирования с очень высоким разрешением. Их вывод заключался в том, что «оптимизированный метод на основе GMM работает, по крайней мере, сравнимо или даже превосходит современные классификаторы, такие как метод опорных векторов (SVM), характеризует типы искусственного земного покрова лучше, чем традиционные методы, смешивает спектральные и текстурные характеристики изображения VHR должным образом, и при этом имеет меньшую вычислительную сложность».

Методы регрессии

Линейная регрессия является одной из наиболее распространенных эмпирических моделей. Здесь переменная отклика используется в своей естественной форме или преобразуется, чтобы быть более симметричной, например, посредством логарифмического преобразования, если распределение очень асимметрично. Затем ответ оценивается линейной комбинацией переменных-предикторов. Эти предикторы могут включать в себя исходные входные переменные, полиномы (для описания нелинейных отношений с откликом), комбинации переменных (для описания взаимодействий).

Линейная регрессия может использоваться новаторскими и расширенными способами для предоставления официальной статистики. Например, Liao (2017 г.) использовал этот метод для определения влияния потребления энергии на температуру поверхности в городских районах. В центре внимания исследования были 32 крупных города Китая. В этом анализе использовалось пять шагов:

1. Получение данных об общем количестве ночных огней и общем потреблении энергии (по провинциям в год).
2. Использование данных НЗ для определения городских районов.
3. Оценка общего потребления энергии в городских районах на основе общего освещения в ночное время.
4. Использование данные НЗ для разработки карт поверхностной температуры.
5. Увязка общего потребления энергии с температурой поверхности.

Масштаб городских территорий:

Источники данных, использованные в этом исследовании, включали:

- Карту растительного покрова/использования земли Китая в 2010 году, созданную путем визуальной интерпретации изображений Landsat TM/ETM + изображений из Центра данных по ресурсам и наукам об окружающей среде Китайской академии наук, которые использовалась для определения шести типов землепользования: искусственные поверхности, обрабатываемые земли, леса, пастбища, водоемы и неиспользуемые земли.
- Температуру поверхности земли (LST), полученную от Aqua MODIS за 8 дней, с пространственным разрешением 1 км, за период с 2008 по 2012 год, в дневное и ночное время.
- Снимки ночного освещения (NTL) Национального управления океанических и атмосферных исследований (NOAA) Национального центра геофизических данных (NGDC) за 2008-2012 гг.

- Общее потребление энергии за каждый год с 2008 по 2012 годы для каждой провинции материкового Китая из Китайского энергетического статистического ежегодника, за исключением Тайваня и Тибета.

Для оценки общего потребления энергии (EC) по сумме ночных огней (NTL) в каждой провинции использовалась модель регрессии, как показано ниже.

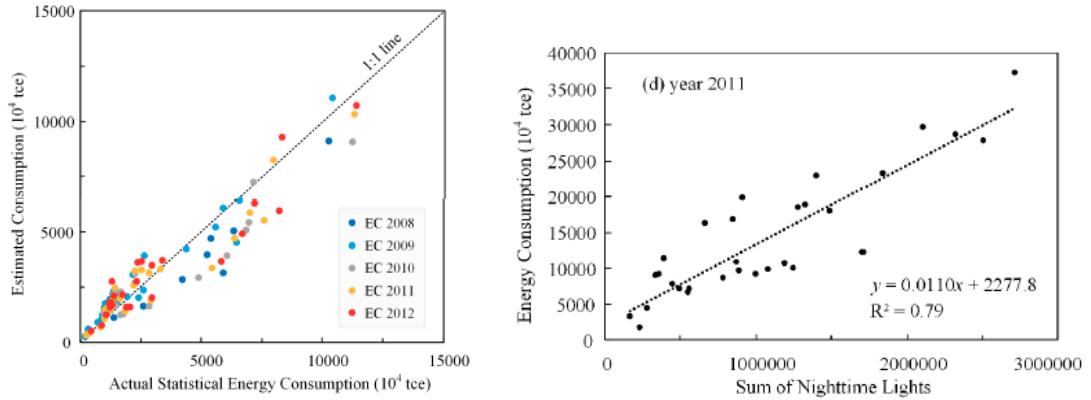


Рисунок 40: Слева: линейный регрессионный анализ для оценки энергопотребления по сумме ночных огней. Справа: сравнение фактического и расчетного энергопотребления. Liao (2017 г.).

Результаты включали оценки пространственных и временных различий в корреляции между городской и пригородной разницей в среднегодовой интенсивности энергопотребления (ΔECI) и SUHI.

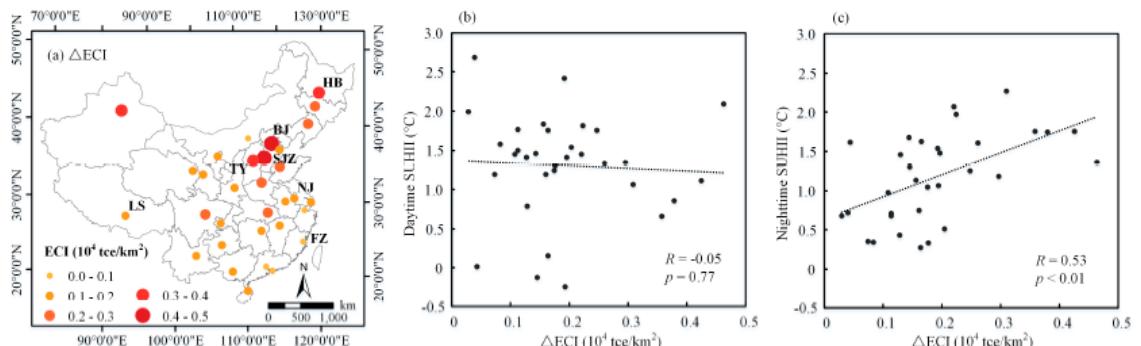


Рисунок 41: Выдержки из работы Liao (2017 г.), на которых показана карта дифференциального энергопотребления между городом и пригородом (слева) и линейная зависимость между этой оценкой и дневным энергопотреблением (в центре) и ночным (справа) SUHI.

Существует множество расширений моделей линейной регрессии для анализа данных НЗ. Три таких примера описаны ниже.

Обобщенные линейные модели (GLM) могут более подходящим образом включать распределение (ненормальной) переменной отклика и гибко связывать ее с линейной функцией независимых (входных) переменных (например, моделирование рассчитывается как распределение Пуассона с логарифмической связью). GLM сами по себе могут быть расширены для обеспечения большей гибкости, например, в виде обобщенных линейных смешанных моделей (GLMM), которые описывают группировку данных. Обзоры использования GLM для данных НЗ предоставлены Morisette & Khorram (1997 г.) и Morisette, Khorram & Mace (1999 г.) в контексте спутникового обнаружения изменений.

Пространственные модели являются расширениями моделей регрессии, которые учитывают тот факт, что географически близкие области с большей вероятностью будут похожи, чем те участки территории, которые значительно удалены друг от друга. Эта пространственная автокорреляция может быть включена в модель для улучшения прогнозов. Аналогичным образом в регрессионной модели могут быть описаны временные закономерности, такие как ежедневные пики и спады, сезонность или более длительно изменяющиеся тренды.

Пространственно-временные модели включают в изображение шаблоны, которые меняются в пространстве или пространстве и времени соответственно. Эти паттерны можно моделировать по-разному, например, как изменяющиеся во времени пространственные признаки, или пространственно изменяющиеся модели в разные периоды времени, или как полное взаимодействие между тремя измерениями.

Другие расширения включают регрессию опорных векторов (SVR), регрессию гребня ядра (KRR), регрессию случайного леса (RFR) и частичную регрессию наименьших квадратов (PLSR). Они описаны Okujeni & van der Linden (2014 г.) в контексте количественной оценки городского земельного покрова по данным НуМар с пространственным разрешением 3,6 и 9 м. Эти авторы обнаружили, что SVR и KRR обеспечивают высокую точность для картирования сложных типов городских поверхностей, т. е. крыш, тротуаров, участков, покрытых травой и деревьями.

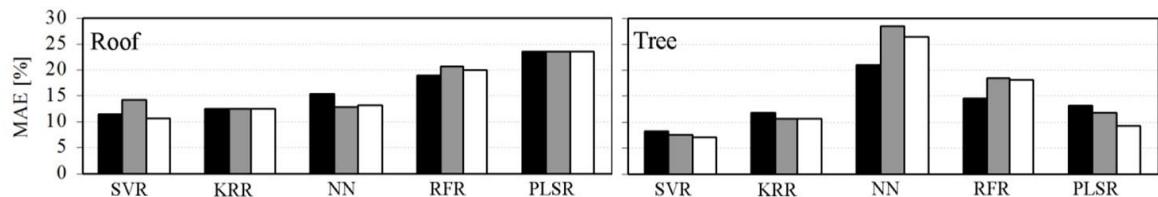


Рисунок 42: Средняя абсолютная ошибка (MAE%) в идентификации объекта (слева Крыша, справа Дерево) из различных регрессионных анализов данных НЗ.

Деревья регрессии представляют собой вариант методов дерева классификации, описанных выше в разделе о подходах к классификации, но здесь ответ является непрерывным, и цель состоит в том, чтобы минимизировать разницу в ответах внутри групп и максимизировать разницу в среднем ответе между группами. Затем дерево будет предсказывать ожидаемый ответ для нового объекта, основываясь на следовании ветвям дерева до конечного узла и вычислении среднего значения ответов в этой группе объектов. Неопределенность прогноза также можно оценить по дисперсии ответов в этой группе. Существует множество различных типов деревьев регрессии, в том числе деревья ускоренной регрессии (BRT) и случайные леса (RF), а также деревья с градиентным усилением (бустингом) (GBM). Например, в GBM деревья регрессии складываются вместе, так что каждое новое дерево обучается на ошибках предыдущих деревьев, сложенных вместе. Сумма этих деревьев затем используется для прогнозирования значения новых точек данных.

Превосходное введение и руководство по бустингу деревьев регрессии с реализацией в статистическом программном пакете R предоставили Elith, Leathwick & Hastie (2008 г.).

Классификация спектральных углов основана на косинусе угла между спектрами отражения пикселей. Он также известен как $\cos\theta$ в других дисциплинах. Более правильно, спектральный угол – это метрика расстояния или сходства, которую можно использовать для оценки сходства с эталонными спектрами с помощью соответствующего правила или критериев принятия решений. Примеры работы с использованием подходов спектрального угла в приложениях дистанционного зондирования включают Sohn & Rebello (2002 г.), Kuching и др. (2007 г.), Petropoulos и др. (2010 г.), Yuhas и др. (1992 г.), Sohn и др. (1999 г.).

Нейронные сети (NN) являются популярными методами регрессии. См. описание в разделе «Классификация» выше. Инновационный пример использования NN для регрессионного анализа с данными НЗ привели Xie и др. (2016 г.). Эти авторы используют подход «переноса обучения»,

который включает в себя последовательное сочетание различных анализов (один из которых – NN) для составления карты бедности в Уганде на основе интенсивности ночных освещения, полученной по спутниковым снимкам. Результаты проиллюстрированы ниже.

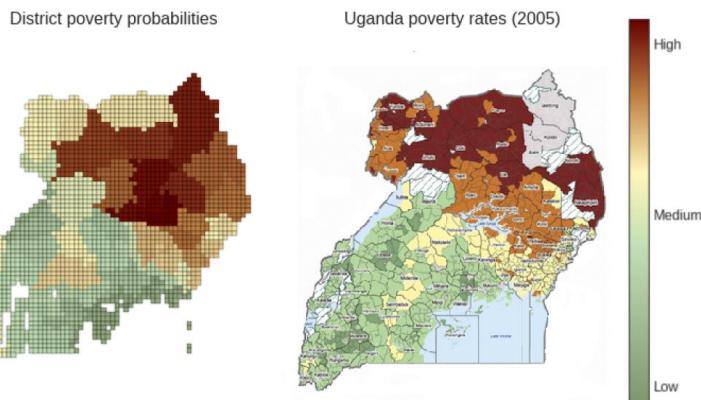


Рисунок 43: Расчетная (слева) и зарегистрированная (справа) вероятность бедности, частично основанная на анализе нейронной сети. Xie и др. (2016 г.)

Методы уменьшения размерности

Анализ основных компонентов (PCA) и его варианты, такие как факторный анализ (FA), являются популярными методами уменьшения размерности эмпирических данных и машинного обучения для анализа данных НЗ. Эти подходы создают линейные комбинации входных переменных, так что новые комбинации (называемые компонентами, факторами, индексами и т. д.) инкапсулируют большой объем информации или изменение данных. Анализ PCA, например, создает независимый набор компонентов, каждый из которых имеет собственное значение («собственный» означает «конкретный»), указывающее долю вариации, объясняемую компонентом, и набор приписанных весов каждой входной переменной (собственному вектору), которая указывает относительную значимость каждой переменной в компоненте.

Методы PCA применялись к данным НЗ уже более десяти лет. Например, Cablk & Minor (2003 г.) использовали четыре основных компонента (PC), полученные из четырех многоспектральных каналов спутника IKONOS с высоким разрешением, в исследовании Саут-Лейк-Тахо, Калифорния, чтобы определить представление непроницаемого водного покрытия в каждом из четырех PC. Это имеет решающее значение для оценки воздействия на качество воды, дикую природу и среду обитания рыб. В другой статье того же времени Call и др. (2003 г.) проверили гипотезу о том, что локальные спектральные сигнатуры вместе с измерениями затухания в водной толще, спутниковыми снимками и географическими информационными системами могут эффективно различать рифовые субстраты. Анализ основных компонентов восходящего излучения показал, что длины волн от 515 до 580 нм наиболее полезны для различия субстратов.

Вот еще три недавних примера. Curzio & Magliulo (2010 г.) оценили геоморфологические последствия эрозии почвы в бассейновом масштабе путем объединения данных Landsat ETM 7+ и геоморфологических данных полевых исследований. Авторы извлекли признаки с помощью PCA на исходных данных и построили из них составные изображения, которые затем использовались в алгоритме классификации. Lawley и др. (2011 г.) применили анализ основных компонентов к спутниковым снимкам за 25 лет, для выявления главных факторов, влияющих на модели роста растительности в засушливых районах, и регионов с аналогичной долгосрочной реакцией. Доминирующими факторами изменчивости были определены пространственное распределение общего роста растительности, сезонность роста, величина сезонной изменчивости роста и регулярность вариации роста. Наконец, Estornell и др. (2013 г.) проанализировали три изображения Landsat, выбранные из двух областей, расположенных в муниципалитетах Гандия и Валлат, оба в провинции Валенсия (Испания). В первом районе исследований использовался один снимок Landsat

от 2005 г. Во втором районе использовались два снимка Landsat (с 1994 по 2000 годы) для выявления существенных изменений земного покрова. Авторы обнаружили, что второй главный компонент изображения района Гандии позволяет обнаружить наличие растительности. Тот же компонент в районе Валлата позволил обнаружить участок лесного хозяйства, пострадавший от лесного пожара.

Хотя РСА является очень полезным и эффективным методом уменьшения размерности при анализе спутниковых изображений, как и у любого метода уменьшения размерности, у него есть недостаток: доминирующие компоненты (факторы) не содержат *всей* информации в данных. Таким образом, обычная практика выбора только доминирующих компонентов и их использования для дальнейшей интерпретации и анализа может иногда приводить к неточной оценке ковариационных матриц, а также к потере пространственной информации. Другие предостережения относительно метода РСА заключаются в том, что он может быть ненадежным, если доступно только небольшое количество изображений, и может потребовать много времени и ресурсов при применении к данным больших изображений. Для преодоления некоторых из этих проблем были предложены различные расширения РСА.

Анализ независимых компонентов (ICA) – это метод, который находит так называемые «независимые компоненты» наблюдаемых данных. Предполагается, что эти данные являются реализациями линейных смесей неизвестных скрытых переменных (независимых компонентов), которые, в свою очередь, предполагаются негауссовскими и взаимно независимыми. ICA разделяет многомерный сигнал на дополнительные подкомпоненты.

Анализ независимых компонентов в определенном смысле связан как с анализом главных компонентов, так и с факторным анализом. Однако одно из преимуществ анализа ICA по сравнению с этими методами заключается в том, что он часто работает в ситуациях, когда другие могут не сработать.

A.2 Краткое изложение методов для статических данных

В следующей таблице приведены некоторые из упомянутых выше методов и приведены некоторые рекомендации по выбору наиболее подходящего метода в определенных ситуациях. Важно понимать, что каждый метод имеет свои преимущества и ограничения, и что в любом конкретном случае применения использование более чем одного метода может дать более глубокое понимание, чем применение любого из них в отдельности.

Техника	Параметрический?	Вычислительная нагрузка (скорость обучения)	Объем требуемой настройки параметров	Вероятности членства в классе?	Пространственные и временные зависимости?	Небольшой объем обучающих данных?	Склонность к переобучению?
SVM	Нет	Медленная	Значительный	Нет	Нет	Да	Нет
Нейронные сети	Нет	Умеренная	Значительный	Нет	Нет	Возможно	Да
Деревья решений	Нет	Быстрая	Низкий	Да	Нет	Нет	Да
Деревья ускоренной регрессии	Нет	Быстрая	Низкий	Да	Нет	Нет	Не точно
Мультиномиальная логистическая регрессия.	Да	Медленная	Умеренный	Да	Могут быть модифицированы	Нет	Нет, под контролем пользователя
K-нн	Нет	Быстрая	Низкий	Нет	Нет	Нет	Да
SSM	Да	Медленная	Умеренный	Зависит от обстоятельств	Да	Нет	Под контролем пользователя
На основе INLA	Да	Быстрее, чем такие эквивалентные варианты, как MCMC	Умеренный	Зависит от обстоятельств	Да	Нет	Под контролем пользователя

A.3 Методы анализа динамических данных

Как описано в главе 3, существует множество эвристических, статистических методов, а также методов машинного обучения для анализа данных НЗ с течением времени.

Сравнительные подходы

Эти подходы достаточно популярны, как описано в Разделе 3. Fulcher и др. (2014 г.) описывают, как два подхода, упомянутые в этом разделе, можно использовать для классификации. Путем создания базы данных об известных временных рядах и их классификациях можно классифицировать новые временные ряды методом сопоставления показателей расстояния между рядами данных. Метод, предложенный Fulcher и др. (2014 г.), изучает свойства данного класса временных рядов и классифицирует новые временные ряды в соответствии с этими изученными свойствами.

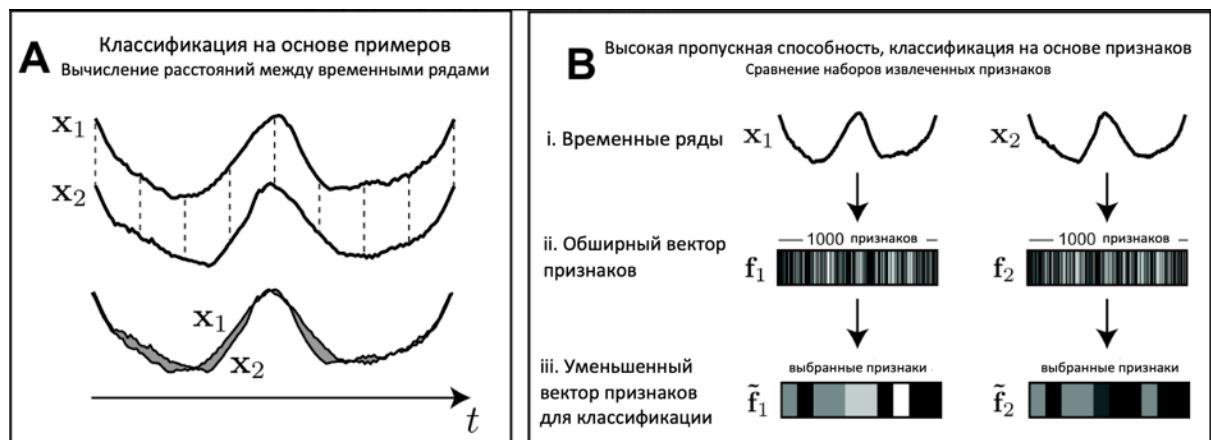


Рисунок 44: Сравнение временных рядов по их структурным характеристикам.

Источник: Fulcher и др. (2014 г.). Какая разница? Различие двух наборов сигналов. <http://systems-signals.blogspot.com.au/>

Кубические сплайны

Эти модели представляют собой полупараметрические уравнения регрессии, которые соответствуют ряду (кубических) полиномов временным данным (Pollock, 2009 г.); см., например, Reed и др. (1994 г.) в качестве примера их использования при анализе фенологических параметров Нормализованного разностного вегетационного индекса (NDVI) с использованием спутниковых изображений.

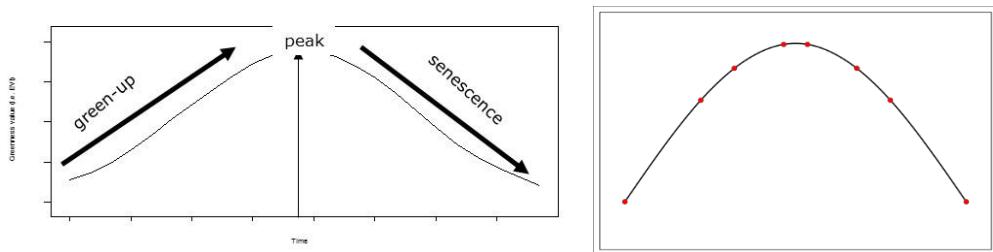


Рисунок 45: Моделирование роста сельскохозяйственных культур (слева) с использованием модели кубического сплайна (справа)

Гармонические модели

Эти модели соответствуют данным более структуризованными синусоидальными (в форме синусоидальной кривой) кривыми. Здесь параметрами модели являются фаза и амплитуда. Например, Potgieter и др. (2007, 2010, 2011 гг.) используют эти модели для оценки урожая с использованием данных НЗ. Расширенный вегетационный индекс (EVI) по 16-дневным спутниковым снимкам MODIS в период посева (т. е. с апреля по ноябрь) был исследован для оценки посевных площадей пшеницы, ячменя, нута и общей посевной площади озимых в регионе на северо-востоке Австралии. Авторы обнаружили значительный потенциал для оценки общей площади озимых культур и высокую точность на пиксельной шкале (>98% верной классификации) для определения общей площади озимых культур. Однако различие самих культур было менее точным.

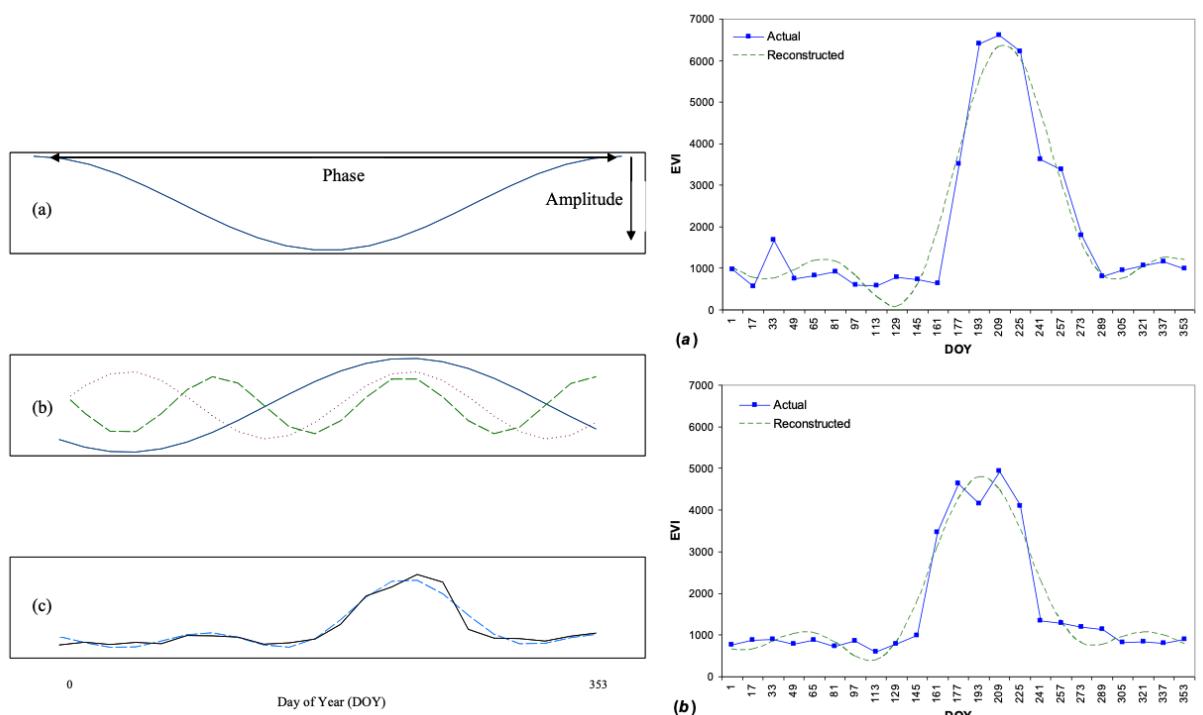


Рисунок 46: Иллюстрация гармонического анализа (слева) и его применения к снимкам MODIS временных рядов озимых культур в 2005 г. (вверху справа) и 2006 г. (внизу справа). Potgeiter и др. (2007, 2011, 2013 гг.)

Спектральная декомпозиция

Подобно тому, как анализ основных компонентов (PCA) и анализ независимых компонентов (ICA) являются формами уменьшения размерности статических данных, методы спектральной декомпозиции, такие как преобразование Фурье, могут использоваться в качестве методов уменьшения размерности для динамических данных. Эти методы извлекают важную информацию из большего пространства в меньшее. Анализ может быть самоцелью, или пространство меньшего размера может быть передано другому статистическому алгоритму или алгоритму машинного обучения для формирования ансамблевого подхода.

Следуя примеру, описанному выше, Potgieter и др. (2007 г.) используют спектральную декомпозицию в форме быстрого преобразования Фурье для данных временных рядов MODIS, результаты которого затем используются в дискриминантном анализе. Аналогичный подход (спектральная декомпозиция временных данных дистанционного зондирования, подаваемых в метод контролируемой классификации дискриминантного анализа) был предпринят Potgieter и др. (2011 г.). Альтернативный подход заключается в подгонке параметрических кривых к данным дистанционного зондирования о росте сельскохозяйственных культур (предыдущие подходы в этом смысле можно отнести к категории непараметрических). Potgieter и др. (2013 г.) используют параметрические функциональные кривые Гаусса для уменьшения размерности данных и снова

классифицируют результирующие параметры кривой с помощью дискриминантного анализа. Не всегда ясно, выполняются ли в этих случаях допущения дискриминантного анализа (например, многомерная нормальность).

Функциональный анализ данных (FDA)

FDA – это непараметрический метод описания и классификации кривых, в котором каждая точка выборки во временном ряду рассматривается как функция, наблюдаемая вдоль основного континуума (например, времени). Это обеспечивает большую гибкость в описании базовой кривой.

Подходы FDA существуют уже более двух десятилетий. Например, метод FDA для логистической регрессии был описан Escabias и др. (2005 г.). В данной работе авторы смоделировали взаимосвязь между риском засухи и кривыми температур. Модель функциональной логистической регрессии также использовалась Cardot и др. (2003 г.) для прогнозирования землепользования на основе временной эволюции данных дистанционного зондирования Земли с грубым разрешением.

Более поздняя статья на эту тему по анализу чередующихся функциональных факторов для улучшения оценки периодических временных тенденций в данных НЗ написана Liu и др. (2012 г.). Эти авторы продемонстрировали свой подход на шестилетних временных рядах с 8-дневными интервалами измерений вегетационного индекса, полученных по изображениям дистанционного зондирования.

Государственные космические модели

Модели пространства состояний были рекомендованы Австралийским бюро статистики (ABS) для анализа динамических данных НЗ. См. статью Tam & Clark (2015 г.) «Статистическая основа для анализа больших данных», в которой рассказывается об использовании SSM со ссылкой на официальную статистику.

Модели пространства состояний являются очень распространенной формой пространственно-временной модели, которая может обеспечить естественную среду для анализа временных рядов данных НЗ, особенно если моделируемый процесс можно рассматривать как процесс, который пространственно развивается во времени. В зависимости от характера данных (геостатистические или сеточные) в модель могут быть включены различные структуры пространственной, временной и пространственно-временной зависимости. Cressie & Wikle (2011 г.) посвящают целую главу (глава 7) различным перестановкам моделей пространственной и временной зависимости, которые могут быть включены в то, что они называют динамическими пространственно-временными моделями (DSTM), и еще одну главу (глава 8) – методам вывода для этих моделей.

На сегодняшний день наиболее распространенным подходом к моделированию пространственно-временных данных является предположение о пространственном процессе, развивающемся во времени (например, Chen и др. (2006 г.), Waller и др. (1997 г.), Lagazio и др. (2001 г.), Knorr-Held & Besag (1998 г.), Mugglin и др. (2002 г.), Jin и др. (2005 г.), Jin и др. (2005 г.), Norton & Niu (2009 г.), Allcroft & Glaseby (2003 г.), Zhu и др. (2005 г.), Zhu и др. (2008 г.), Zheng & Zhu (2008 г.), Daniels и др. (2006 г.), Ugarte и др. (2010 г.)). Как отмечено у Cressie & Wikle (2011 г.), этот подход концептуально целесообразен во многих случаях применения и позволяет пространственно-временному процессу соответствовать подходу динамического пространственного моделирования или структуре моделирования пространства состояний. Этот подход был изучен Австралийским бюро статистики в качестве пилотного проекта.

Линейная гауссовская модель пространства состояний принимает общий вид:

$$\begin{aligned} Y_t &= H_t X_t + v_t, & v_t &\sim N_{m_t}(0, R_t) \\ X_t &= M_t X_{t-1} + w_t, & w_t &\sim N_n(0, Q_t), \end{aligned}$$

где для времени t , $t = 1, \dots, T$, Y_t обозначает m_t – вектор размерных данных, X_t обозначает n – мерный ненаблюдаемый вектор «состояния», v_t – серийно независимая ошибка наблюдения, а w_t – серийно независимая инновационная ошибка (Shumway & Stoffer, 2011 г.). Обычно считается, что v_t и w_t являются взаимно независимыми. Также обычно предполагается, что матрица наблюдения H_t , матрица переходов состояний M_t (также известная как матрица эволюции или пропагатора) и две ковариационные матрицы R_t и Q_t известны, хотя это требование может быть ослаблено.

Обычно основным механизмом учета пространственно-временных зависимостей являются ковариационные матрицы R_t и Q_t . Например, обычно можно просмотреть состояния X_t как движимые не только временными, но и пространственными процессами, поэтому на Q_t накладывается некоторая форма пространственной структуры. Это может быть одна из распространенных структур функции ковариации для геостатистических данных или, возможно, разреженная ковариационная матрица на основе соседних данных, индуцированная предположением о Марковском случайному поле в сеточных данных. Пространственные, временные или пространственно-временные структуры также могут быть допущены в структуре измерения.

Расширения этих моделей включают негауссовые модели в пространстве состояний, нелинейные модели уравнений наблюдения, агентно-ориентированные подходы, обобщенные аддитивные модели смешанных эффектов и многомерные модели отклика. Другой важной областью исследований в этом контексте является уменьшение размерности с помощью спектральных декомпозиций. Такое уменьшение размерности может существенно снизить вычислительную нагрузку при подгонке этих динамических моделей к данным.

Модели пространства состояний использовались для оценки и прогнозирования урожая более 15 лет. Например, Wendroth и др. (2003 г.) использовали этот подход для прогнозирования урожайности ячменя. Совсем недавно Schneibel и др. (2017 г.) применили этот метод к ежегодным данным временных рядов Landsat с целью выявления процессов деградации сухих лесов в южно-центральной части Анголы.

Пример использования скрытой Марковской модели для анализа данных НЗ привели Yuan и др. (2015 г.) в контексте обнаружения посягательств городов на сельскохозяйственные угодья в Китае. В этом исследовании были использованы два источника данных:

- Данные MODIS за 2001–2010 гг. (крупномасштабный охват, высокое временное разрешение [16 дней]), загруженные с веб-сайта Системы архивирования и распространения данных первого уровня обработки и атмосферных данных (LAADS); это дало четыре канала спектрального отражения для растительности и поверхности земли, т. е. канал 1 (красный: 620–670 нм), канал 2 (NIR: 841–875 нм), канал 3 (синий: 459–479 нм), канал 7 (MIR: 2105–2155 нм)
- Карта ESA Global Land Cover (GlobCover) за 2009 год, для обучения модели: сельскохозяйственные угодья, застроенные классы.

В качестве этапа предварительной обработки был использован функциональный анализ (посредством модели Фурье, см. выше) для создания высококачественных временных рядов коэффициентов отражения. Анализ k-средних также применялся для кластеризации землепользования с использованием обучающих данных. Затем для сглаживания оценок использовался метод скользящего окна, а для выявления изменений в классах применялась Скрытая Марковская модель. Авторам удалось обнаружить 96% правильных изменений и 0,04% ложных сигналов тревоги.

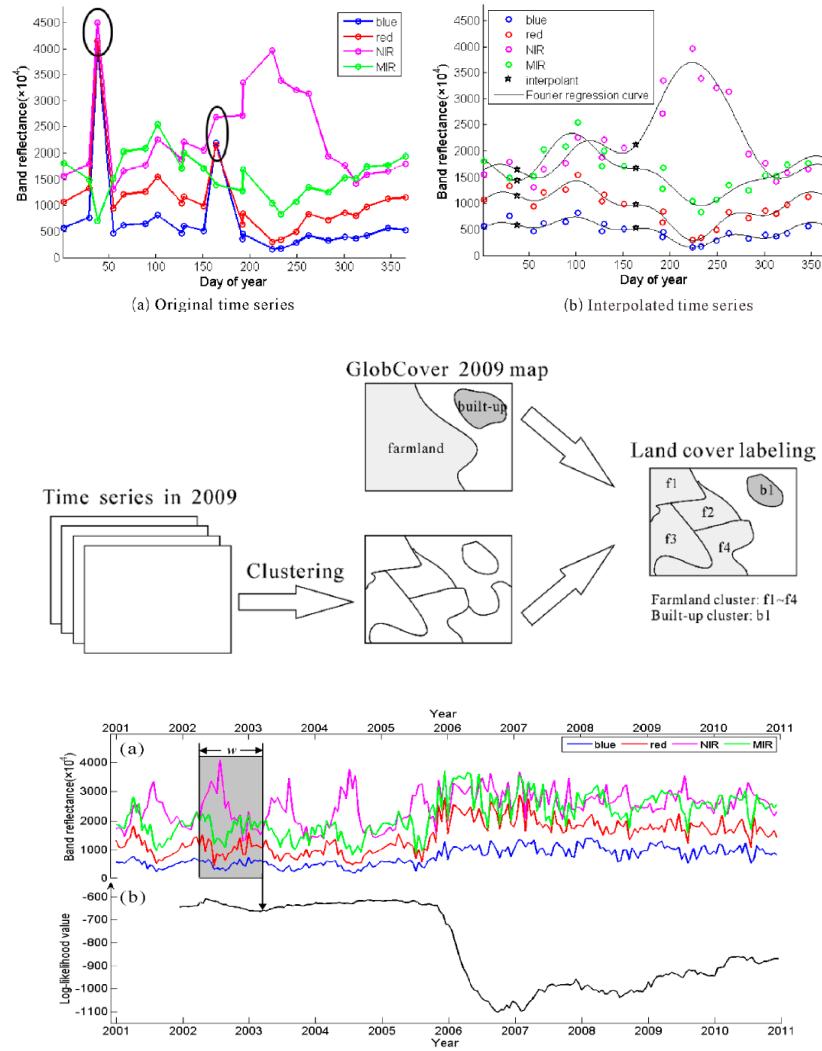


Рисунок 47: (Верху) Функциональный анализ, используемый для интерполяции временных рядов данных НЗ. (В центре) Кластеризация k-средних, используемая для классификации сельскохозяйственных угодий. (Внизу) Скрытая Марковская модель для выявления роста городов. Yuan и др. (2015 г.)

A.4 Выбор между методами

Сравнения между различными эмпирическими методами и подходами машинного обучения проводились более десяти лет. Преимущества и недостатки различных методов зависят от характера проблемы. Некоторые примеры этих сравнений приведены ниже:

- Сравнение между методом машинного обучения (ML) и полиномиальной (или политомической) логистической регрессией для анализа данных дистанционного зондирования провели Hogland и др. (2013 г.). Их вывод заключался в том, что методы логистической регрессии были менее строгими, более гибкими и простыми для интерпретации. Авторы также сослались на документы, которые показывают, что эти подходы работают не хуже других методов машинного обучения; см. ниже.
- De Melo и др. (2003 г.) сравнили ML со смешанными гауссовскими моделями для контролируемой классификации мультиспектральных изображений дистанционного зондирования, отдавая предпочтение смешанным гауссовским моделям.
- Shao & Lunetta (2012 г.) сравнили метод опорных векторов (SVM) с нейронными сетями и CART (деревьями классификации и регрессии) на ограниченных обучающих данных для классификации земного покрова с использованием данных временных рядов MODIS.

- Otukei & Blaschke (2010 г.) рассмотрели оценку изменения земного покрова с использованием деревьев решений, SVM (с функцией ядра радиального базиса) и классификатора максимального правдоподобия и пришли к выводу, что деревья решений работают лучше всего.
- Szuster и др. (2011 г.) сравнили классификацию максимального правдоподобия и ANN с SVM для землепользования и классификации земного покрова в тропических прибрежных зонах.
- Yang и др. (2011 г.) оценили различные контролируемые классификаторы, используя изображения SPOT 5 (спутник с высоким разрешением), и обнаружили, что методы максимального правдоподобия и метод опорных векторов работают лучше всех рассмотренных.
- Huang и др. (2002 г.) исследовали SVM для классификации земного покрова и сравнивали их с нейронными сетями, методами максимального правдоподобия и деревьями решений.
- Melgani и Bruzzone (2004 г.) рассмотрели SVM для данных дистанционного зондирования, сравнив их использование с нейронными сетями (радиальные базисные функции NN) и K-ближайшим соседом (K-NN). Они пришли к выводу, что SVM являются жизнеспособным вариантом.

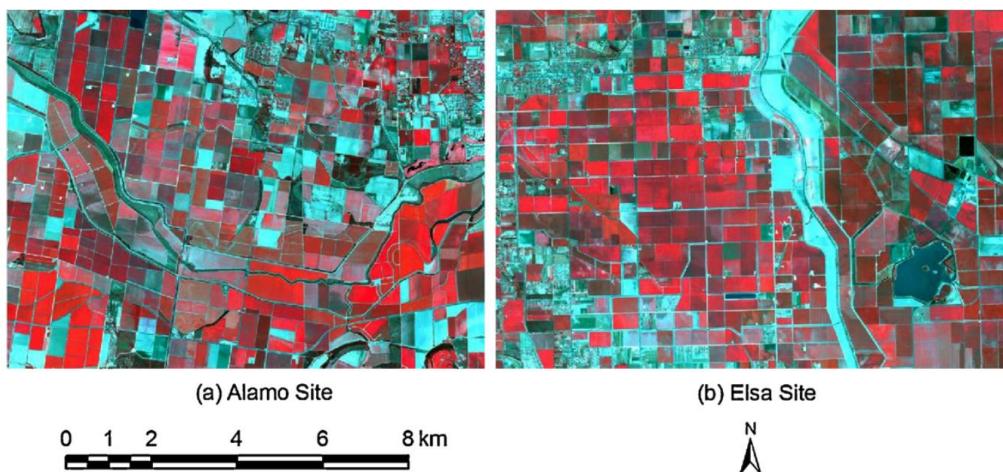


Рисунок 48: Идентификация культур с использованием данных Н3. Yang и др. (2011 г.)

A.5 Комбинированные методы

Вместо выбора одного метода все более популярным подходом в литературе по машинному обучению является их комбинирование.

Пример комбинации различных методов для улучшения классификации приводится Sweeney и др. (2015 г.). Эти авторы объединили вероятностное моделирование в форме логистической регрессии с традиционными подходами дистанционного зондирования для получения карт мелкомасштабных пахотных земель. Хотя различные методы хорошо зарекомендовали себя, авторы утверждают, что новизна их подхода заключается в последовательности применения и способе их сочетания.

Пример комбинации различных методов для улучшения кластеризации изображений дистанционного зондирования также приводят Neagoe & Chirila-Berbentea (2016 г.). В этой статье авторы предлагают объединить модели kNN и гауссовой смеси, при этом первый метод используется для определения отправных точек для второго метода, а EM (максимизация ожидания) используется для анализа.

Список литературы

- ABARES, (2016). Agricultural Commodity Statistics 2016. Australian Bureau of Agricultural and Resource Economics and Sciences, Canberra. Retrieved from www.agriculture.gov.au/abares/publications/display?url=http://143.188.17.20/anrdl/DAFFService/display.php%3Ffid%3Dpb_agcstd9abcc0022016_Sn9Dg.xml
- Adams, J. B., Sabol, D. E., Kapos, V., Almeida Filho, R., Roberts, D. A., Smith, M. O. and Gillespie, A. R. (1995). Classification of multispectral images based on fractions of endmembers: Application to land-cover change in the Brazilian Amazon. *Remote sensing of Environment*, 52(2), 137–154.
- Allcroft, D. J. and Glaseby, C. A. (2003). A latent Gaussian Markov random field model for spatio-temporal rainfall disaggregation. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 52(4), 487–498.
- Al-Obeidat, F., Al-Taani, A. T., Belacel, N., Feltrin, L. and Banerjee, N. (2015). A Fuzzy Decision Tree for Processing Satellite Images and Landsat Data. *Procedia Computer Science*, 52, 1192–1197.
- Anderson, K., Ryan, B., Sonntag, W., Argyro, K., Friedl, L. (2017) Earth observation in service of the 2030 Agenda for Sustainable Development, *Geo-spatial Information Science*, 20:2, 77-96.
- Aplin, P. and Atkinson, P. M. (2001). Sub-pixel land cover mapping for per-field classification. *International Journal of Remote Sensing*, 22(14), 2853–2858.
- Aplin, P., Atkinson, P. M., Curran, P. J. and Others (1999). Per-field classification of land use using the forthcoming very fine spatial resolution satellite sensors: problems and potential solutions. *Advances in remote sensing and GIS analysis*, pp. 219–239.
- Atkinson, P. M. and Tatnall, A. R. L. (1997). Introduction neural networks in remote sensing. *International Journal of remote sensing*, 18(4), 699–709.
- Augustin, N. H., Musio, M., von Wilpert, K., Kublin, E., Wood, S. N. and Schumacher, (2009). Modelling spatio temporal forest health monitoring data. *Journal of the American Statistical Association*, 104 (487), 899-911.
- Bai' llo, A., Cuesta-Albertos, J. A. and Cuevas, A. (2011a). Supervised Classification for a Family of Gaussian Functional Models. *Scandinavian Journal of Statistics*, 38, 480– 498.
- Bai' llo, A., Cuevas, A. and Fraiman, R. (2011b). Classification Methods for Functional Data. In *The Oxford Handbook of Functional Data Analysis*. (Eds F. Ferraty and Y. Romain), pp. 259–297. Oxford: Oxford University Press, UK.
- Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. Crc Press, second edition.
- Banerjee, S., Gelfand, A. E., Finley, A. O. and Sang, H. (2008). Gaussian Predictive Process Models for Large Spatial Data Sets. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 70(4), 825–848.
- Bastin, L. (1997). Comparison of fuzzy c-means classification, linear mixture modelling and MLC probabilities as tools for unmixing coarse pixels. *International Journal of Remote Sensing*, 18(17), 3629–3648.

Bateson, C., Asner, G. P., Wessman, C. and Others (2000). Endmember bundles: A new approach to incorporating endmember variability into spectral mixture analysis. *Geo- science and Remote Sensing, IEEE Transactions on*, 38(2), 1083–1094.

Bavaghar, P. (2015) Deforestation modelling using logistic regression and GIS. *Journal of Forest Science* 61, 193-199.

Bédard, F. and Reichert, G. (2013). Integrated Crop Yield and Production Forecasting using Remote Sensing and Agri-Climatic data. Analytical Projects Initiatives final report. Remote Sensing and Geospatial Analysis, Agriculture Division, Statistics Canada

Benediktsson, J. A. and Swain, P. H. (1992). Consensus theoretic classification methods. *IEEE Trans. Syst., Man, Cybern.*, vol. 22, 688–704.

Benediktsson, J. A., Sveinsson, J. R. and Swain, P. H. (1997). Hybrid consensus theoretic classification. *Geoscience and Remote Sensing, IEEE Transactions on*, 35(4), 833–843.

Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I. and Heynen, M. (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of photogrammetry and remote sensing*, 58(3), 239–258.

Berliner, L. M., Wikle, C. K. and Cressie, N. (2000). Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling. *Journal of Climate*, 13(22), 3953–3968.

Berliner, L. M., Wikle, C. K. and Milliff, R. F. (1999). Multiresolution wavelet analyses in hierarchical Bayesian turbulence models. In *Bayesian inference in wavelet-based models* (Eds P. Mueller and B. Vidakovic), Chapter 21, pp. 341–359. New York: Springer.

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M. and Songini, M. (1995). Bayesian analysis of space-time variation in disease risk. *Statistics in medicine*, 14(21-22), 2433–2443.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2), 192–236.

Bigdeli, B., Samadzadegan, F. and Reinartz, P. (2015). Fusion of hyperspectral and LIDAR data using decision template-based fuzzy multiple classifier system. *International Journal of Applied Earth Observation and Geoinformation*, 38, 309–320.

Billheimer, D., Cardoso, T., Freeman, E., Guttorp, P., Ko, H. W. and Silkey, M. (1997). Natural variability of benthic species composition in the Delaware Bay. *Environmental and Ecological Statistics*, 4(2), 95–115.

Blangiardo, M., Cameletti, M., Baio, G. and Rue, H. H. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-temporal Epidemiology*, 7, 39–55.

Blanzieri, E. and Melgani, F. (2008). Nearest neighbor classification of remote sensing images with the maximal margin principle. *Geoscience and Remote Sensing, IEEE Transactions on*, 46(6), 1804–1811.

Bolin, D. and Lindgren, F. (2013). A comparison between Markov approximations and other methods for large spatial data sets. *Computational Statistics & Data Analysis*, 61, 7–21.

Bradley, J. R., Wikle, C. K. and Holan, S. H. (2016) Bayesian Spatial Change of Support for Count-Valued Survey Data with Application to the American Community Survey. *Journal of the American Statistical Association*, 0(ja), 1–43.

- Breiman, L. (1996). Bagging Predictors. *Machine Learning*, 24(2), 123–140.
- Breiman, L. (2001) Statistical modeling: the two cultures. *Statistical Science* 16(9), 199-231.
- Breiman, L., Friedman, J., Stone, C. J. and Olshen, R. A. (1984). *Classification and Regression Trees*. CRC press.
- Briem, G. J., Benediktsson, J. A. and Sveinsson, J. R. (2002). Multiple classifiers applied to multisource remote sensing data. *Geoscience and Remote Sensing, IEEE Transactions on*, 40(10), 2291–2299.
- Brynjarsdottir, J. and Berliner, M. (2014). Dimension–Reduced Modelling of Spatio–Temporal Processes. *Journal of the American Statistical Association*, 109(508), 1647– 1659.
- Cablk, M. E. and Minor, T. B. (2003). Detecting and discriminating impervious cover with high-resolution IKONOS data using principal component analysis and morphological operators. *International Journal of Remote Sensing*, 24: 4627–4645.
- Calis, N. and Erol, H. (2012). A new per-field classification method using mixture discriminant analysis. *Journal of Applied Statistics*, 39(10), 2129–2140.
- Calder, C. A. (2007). Dynamic factor process convolution models for multivariate space- time data with application to air quality assessment. *Environmental and Ecological Statistics*, 14(3), 229–247.
- Call, K.A., Hardy, J. T., and Wallin, D. O. (2003). Coral reef habitat discrimination using multivariate spectral analysis and satellite remote sensing. *International Journal of Remote Sensing*, 24: 2627– 2639.
- Cao, Q., Guo, Z. and Yang, Y. (2015). An improved back propagation neural network approach to the remote sensing land use and land cover classification. *Computer Science and Applications: Proceedings of the 2014 Asia-Pacific Conference on Computer Science and Applications (CSAC 2014)*, Shanghai, China, 27-28 December 2014, 369. CRC Press.
- Carlin, B. P., Polson, N. G. and Stoffer, D. S. (1992). A Monte Carlo approach to non-normal and nonlinear state-space modelling. *Journal of the American Statistical Association*, 87(418), 493–500. Carr, J. R. (1999). Classification of digital image texture using variograms. *Advances in remote sensing and GIS analysis*, pp. 135–146.
- Cavalli, R.M., Licciardi, G.A., and Chanussot, J. (2013). Detection of Anomalies Produced by Buried Archaeological Structures Using Nonlinear Principal Component Analysis Applied to Airborne Hyperspectral Image. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6: 659 – 669.
- Cardot H, Faivre R, Goulard M (2003) Functional approaches for predicting land use with the temporal evolution of coarse resolution remote sensing data. *Journal of Applied Statistics* 30: 1185-1199.
- Chasmer, L., Hopkinson, C., Veness, T., Quinton, W. and Baltzer, J. (2014). A decision- tree classification for low-lying complex land cover types within the zone of discontinuous permafrost. *Remote Sensing of Environment*, 143, 73–84.
- Chen, K. S., Tzeng, Y. C., Chen, C. F. and Kao, W. L. (1995). Land-cover classification of multispectral imagery using a dynamic learning neural network. *Photogrammetric Engineering and Remote Sensing*, 61(4), 403–408.
- Chen, L., Fuentes, M. and Davis, J. M. (2006). Spatial-temporal statistical modelling and prediction of environmental processes. In *Hierarchical Modelling for the Environmental Sciences: Statistical methods*

and applications (Eds J. S. Clark and A. E. Gelfand), chapter 7, pp. 121–144. New York: Oxford University Press.

Chiang, S.-S., Chang, C.-I. and Ginsberg, I. W. (2000). Unsupervised hyperspectral image analysis using independent component analysis. *Geoscience and Remote Sensing Symposium Proceedings. IGARSS 2000. IEEE 2000 International*, volume 7, pp. 3136–3138.

Chipanshi, A., Zhang, Y., Kouadio, L., Newlands, N., Davidson, A., Hill, H., Warren, R., Qian, B., Daneshfar, B., Bedard, F. and Reichert, G. (2015). Evaluation of the Integrated Canadian Crop Yield Forecaster (ICCYF) Model for In-season Prediction of Crop Yield across the Canadian Agricultural Landscape. *Agricultural and Forest Meteorology*, 206:137–150. DOI: <http://dx.doi.org/10.1016/j.agrformet.2015.03.007>

Clark, J. S. and Gelfand, A. E. (2006). *Hierarchical Modelling for the Environmental Sciences: Statistical methods and applications*. New York: Oxford University Press, UK.

Clayton, D. G. (1996). Generalized Linear Mixed Models. In *Markov Chain Monte Carlo in Practice* (Eds W. Gilks, S. Richardson and D. Spiegelhalter), 275–301. Chapman & Hall.

Clayton, D. G. and Bernardinelli, L. (1992). Bayesian methods for mapping disease risk. In *Geographical and Environmental Epidemiology: Methods for Small-Area Studies* (Eds P. Elliot, J. Cuzick, D. English and R. Stern), 205–220. UK: Oxford University Press.

Cliff, A. D. and Ord, J. K. (1981). *Spatial processes: models and applications*. London, UK: Pion.

Clinton, N., Yu, L. and Gong, P. (2015). Geographic stacking: Decision fusion to increase global land cover map accuracy. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103, 57–65.

Collins, M. J., Dymond, C. and Johnson, E. A. (2004). Mapping subalpine forest types using networks of nearest neighbour classifiers. *International Journal of Remote Sensing*, 25(9), 1701–1721.

Cowles, M. K. (2003). Efficient model-fitting and model-comparison for high-dimensional Bayesian geostatistical models. *Journal of Statistical Planning and Inference*, 112(1-2), 221–239.

Crammer, K. and Singer, Y. (2002). On the algorithmic implementation of multiclass kernel-based vector machines. *The Journal of Machine Learning Research*, 2, 265–292.

Cressie, N. A. C. (1993). *Statistics for Spatial Data Revised Edition*. Wiley Series in probability and mathematical statistics. New York: John Wiley & Sons, Inc.

Cressie, N. and Huang, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, 94(448), 1330–1339.

Cressie, N. and Verzelen, N. (2008). Conditional-mean least-squares fitting of Gaussian Markov random fields to Gaussian fields. *Computational Statistics and Data Analysis*, 52(5), 2794–2807.

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. Wiley.

Cressie, N., Shi, T. and Kang, E. L. (2010a). Fixed Rank Filtering for Spatio-Temporal Data. *Journal of Computational and Graphical Statistics*, 19(3), 724–745.

Cressie, N., Shi, T. and Kang, E. L. (2010b). Fixed rank filtering for spatio-temporal data.

Cuevas, A. (2014). Journal of Statistical Planning and Inference A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference*, 147, 1–23.

Curzio, S.L. and Magliulo, P. (2010). Soil erosion assessment using geomorphological remote sensing techniques: an example from southern Italy. *Earth Surface Processes and Landforms* 35: 262–271.

Daniels, M. J., Zhou, Z. and Zou, H. (2006). Conditionally Specified Space-Time Models for Multivariate Processes. *Journal of Computational and Graphical Statistics*, 15(1), 157–177.

De Fries, R. S., Hansen, M., Townshend, J. R. G. and Sohlberg, R. (1998). Global land cover classifications at 8 km spatial resolution: the use of training data derived from Landsat imagery in decision tree classifiers. *International Journal of Remote Sensing*, 19(16), 3141–3168.

De Melo, A.C.O, de Moraes, R.M., dos Santos Machado, L. (2003) Gaussian mixture models for supervised classification of remote sensing multispectral images. In *Progress in Pattern Recognition, Speech and Image Analysis*, Volume 2905 of the series *Lecture Notes in Computer Science*, pp. 440- 447.

De Wit, A. J. W. and Clevers, J. (2004). Efficiency and accuracy of per-field classification for operational crop mapping. *International journal of remote sensing*, 25(20), 4091–4112.

Dekker, A., Peters, S., Vos, R., Rijkeboer, M. (2013) Remote sensing for inland water quality detection and monitoring: state-of-the-art application in Friesland waters. In van Dijk, A., Bos, G. (Eds) *GIS and Remote Sensing Techniques in Land and Water management*. CRC Press. Chapter 3.

Delaigle, A. and Hall, P. (2012a). Achieving near perfect classification for functional data. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 74(2), 267–286.

Delaigle, A. and Hall, P. (2012b). Methodology and theory for partial least squares applied to functional data. *Annals of Statistics*, 40(1), 322–352.

Delaigle, A., Hall, P. and Bathia, N. (2012). Componentwise classification and clustering of functional data. *Biometrika*, 99(2), 299–313.

Delsol, L., Ferraty, F. and Martinez-Calvo, A. (2011). Functional Data Analysis: An Interdisciplinary Statistical Topic. In *Statistical Learning and Data Science*. (Eds.)

Dennison, P. E. and Roberts, D. A. (2003). Endmember selection for multiple endmember spectral mixture analysis using endmember average RMSE. *Remote Sensing of Environment*, 87(2), 123–135.

Diggle, P. J. and Ribeiro, P. J. J. (2007). *Model-Based Geostatistics*. New York: Springer.

Doucet, A., de Freitas, N. and Gordon, N. (eds) (2001). *Sequential Monte Carlo methods in practice*. New York: Springer Science & Business Media.

Du, P., Xia, J., Zhang, W., Tan, K., Liu, Y. and Liu, S. (2012). Multiple classifier system for remote sensing image classification: A review. *Sensors*, 12(4), 4764–4792.

Elith, J., Leathwick, J.R., Hastie, T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology* 77, 802-813.

Erbek, F. S., Ozkan, C. and Taberner, M. (2004). Comparison of maximum likelihood classification method with supervised artificial neural network algorithms for land use activities. *International Journal of Remote Sensing*, 25(9), 1733–1748.

Erol, H. and Akdeniz, F. (2005). A per-field classification method based on mixture distribution models and an application to Landsat Thematic Mapper data. *International Journal of Remote Sensing*, 26(6), 1229–1244. Evensen, G. (2009) Data assimilation: the ensemble Kalman filter. New York: Springer Science & Business Media, second edition.

Escabias M., Aguilera, A.M., Valderrama, M.J. (2005). Modelling environmental data by functional principal component logistic regression. *Environmetrics* 16: 95-107.

Esch, T., Taubenbock, H., Roth, A., Heldens, W., Felbier, A., Thiel, M., Schmidt, M., Muller, A., Dech, S. (2012). TanDEM-X mission – new perspectives for the inventory and monitoring of global settlement patterns. *Journal of Applied Remote Sensing* 6(1), 061702-1-061702-21.

Estornell, J., Martí-Gavila, J. M., Sebastia, M. T. and Mengual, J. (2013). Principal component analysis applied to remote sensing. *Modelling in Science Education and Learning*, 6(2): 83–89.

European Commission (2016). MARS bulletins May 2016. Retrieved from <https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/jrc-mars-bulletin-crop-monitoring-europe-may-2016-fairly-good-outlook-despite-unsettled>

Fan, J., Han, F. and Liu, H. (2014). Challenges of Big Data analysis. *National Science Review* (June 2014) 1 (2): 293-314. Retrieved from <http://nsr.oxfordjournals.org/content/1/2/293.full>.

Febrero-Bande, M. and de la Fuente, M. O. (2012). Statistical Computing in Functional Data Analysis: The R Package fda.usc. *Journal of Statistical Software*, 51(4), 1–28.

Fernández-Prieto, D. (2002). An iterative approach to partially supervised classification problems. *International Journal of Remote Sensing*, 23(18), 3887–3892.

Ferraty, F. and Romain, Y. (eds) (2011). *The Oxford Handbook of Functional Data Analysis*. Oxford: Oxford University Press, UK.

Ferreira, M. A. R., Holan, S. H. and Bertolde, A. I. (2011). Dynamic multiscale spatiotemporal models for Gaussian areal data. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 73(5), 663–688.

Finkenstadt, B., Held, L. and Isham, V. (eds) (2007). *Statistical methods for spatio-temporal systems*. Boca Raton, FL: CRC Press.

Fix, E. and Hodges Jr, J. L. (1951). Discriminatory analysis-nonparametric discrimination: consistency properties. Technical report, DTIC Document.

Food and Agriculture Organisation (FAO) (2016). FAO Cereal Supply and Demand Brief. Retrieved from <http://www.fao.org/worldfoodsituation/csdb/en/>.

Foody, G. M. (1996a). Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data. *International Journal of Remote Sensing*, 17(7), 1317– 1340.

Foody, G. M. (1996b). Fuzzy modelling of vegetation from remotely sensed imagery. *Ecological modelling*, 85(1), 3–12.

Foody, G. M. (1997). Fully fuzzy supervised classification of land cover from remotely sensed imagery with an artificial neural network. *Neural Computing & Applications*, 5(4), 238–247.

Foody, G. M. (1999). Image classification with a neural network: from completely-crisp to fully-fuzzy situations. *Advances in remote sensing and GIS analysis*, pp. 17–37.

Foody, G. M. and Arora, M. K. (1997). An evaluation of some factors affecting the accuracy of classification by an artificial neural network. *International Journal of Remote Sensing*, 18(4), 799– 810.

Foody, G. M., Lucas, R. M., Curran, P. J. and Honzak, M. (1997). Non-linear mixture modelling without end-members using an artificial neural network. *International Journal of Remote Sensing*, 18(4), 937–953.

Foody, G. M., McCulloch, M. B. and Yates, W. B. (1995). Classification of remotely sensed data by an artificial neural network: issues related to training data characteristics. *Photogrammetric Engineering and Remote Sensing*, 61(4), 391–401.

Franco-Lopez, H., Ek, A. R. and Bauer, M. E. (2001). Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. *Remote sensing of environment*, 77(3), 251–274.

Freund, Y., Schapire, R. E. and Others (1996). Experiments with a new boosting algorithm. In *ICML*, volume 96, pp. 148–156.

Friedl, M. A. and Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote sensing of environment*, 61(3), 399–409.

Friedman, J., Hastie, T. and Tibshirani, R. (2010) Regularisation Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1), 1–22.

Friedman, J., Hastie, T., Tibshirani, R. and Others (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics*, 28(2), 337–407.

Fruhwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. New York: Springer.

Fulcher, B.D., Little, M.A., Jones, N.S. (2013). Highly comparative time-series analysis: the empirical structure of time series and their methods. *Journal of the Royal Society* <https://doi.org/10.1098/rsif.2013.0048>

Fuentes, M., Chen, L. and Davis, J. M. (2008). A class of nonseparable and nonstationary spatial temporal covariance functions. *Environmetrics*, 19, 487–507.

Gallego, J.C.M., Michaelsen, J., Bossyns, B., Fritz, S. (2008) Best practices for crop area estimation with Remote Sensing. GEOSS Community of Practice Ag 0703a. JRC Scientific and Technical Reports.

Retrieved from

https://www.earthobservations.org/documents/cop/ag_gams/GEOSS%20best%20practices%20area%20estimation%20final.pdf

Gao, F., Masek, J., Schwaller, M., and Hall, F. (2006). On the blending of the Landsat and MODIS surface reflectance: predict daily Landsat surface reflectance. *Remote Sensing*, 44(8), 2207-2218

Gao, F., Anderson, M.C., Zhang, X., Yang, Z., Alfieri, J.G., Kustas, W.P., Mueller, R., Johnson, D.M., and J.H. Prueger. (2017). Toward mapping crop progress at field scales using Landsat and MODIS imagery. *Remote Sens. Environ.*, 188, 9–25.

Geib, C., Taubenbock, H. (2013) Remote sensing contributing to assess earthquake risk: from a literature review towards a roadmap. *Natural Hazards* 68 (1), 7-48.

Gelfand, A. E. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, 4(1), 11–25.

Gelfand, A. E., Diggle, P. J., Fuentes, M. and Guttorp, P. (eds) (2010). *Hand book of Spatial Statistics*. Boca Raton, FL: Chapman & Hall.

Gelfand, A. E., Schmidt, A. M., Banerjee, S. and Sirmans, C. F. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, 13(2), 263–312.

Geneletti, D. and Gorte, B. G. H. (2003). A method for object-oriented land cover classification combining Landsat TM data and aerial photographs. *International Journal of Remote Sensing*, 24(6), 1273–1286.

Gettler-Summa, M., Bottou, L., Goldfarb, B., Murtagh, F., Pardoux, C., and Touati, M. (2011). Statistical Learning and Data Science p. 189–195. Boca Raton, FL: Chapman & Hall.

Ghosh, A., Mishra, N. S. and Ghosh, S. (2011). Fuzzy clustering algorithms for unsupervised change detection in remote sensing images. *Information Sciences*, 181(4), 699–715.

Ghosh, A., Subudhi, B. N. and Bruzzone, L. (2013). Integration of Gibbs Markov random field and Hopfield-type neural networks for unsupervised change detection in remotely sensed multi-temporal images. *Image Processing, IEEE Transactions on*, 22(8), 3087–3096.

Global Strategy 2017. Handbook on Remote Sensing for Agricultural statistics. Rome, www.gsars.org, 250pp.

Gneiting, T. (2002). Nonseparable, Stationary Covariance Functions for Space-Time Data. *Journal of the American Statistical Association*, 97(458), 590–600.

Goldfarb, B., Murtagh, F., Pardoux, C., and Touati, M., pp. 197–203. Boca Raton, FL: Chapman & Hall.

González-Manteiga, W. and Vieu, P. (2011). Methodological Richness of Functional Data Analysis. In *Statistical Learning and Data Science*.

Gordon, N., Salmond, D. and Smith, A. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F Radar and Signal Processing*, 140(2), 107–113.

Grenzdörffer, G. (2001). Requirements and possibilities of remote sensing for precision agriculture - current status and future developments. In: Grenier, G., and Blackmoore S. [ed.] (2001). Proceedings of the Third European Conference on Precision Agriculture, Montpellier June 18th-21st, 1, 211–216.

Guerschman, J. P., Donohue, R. J., Van Niel T. G., Renzullo L. J., Dekker A. G., Malthus T. J., McVicar T. R. and Van Dijk, A. I. J. M. (2016). Earth observation for water resources management Part II in Earth observation for water resources management: current use and future opportunities for the water sector, (2016) Eds. García, L., Rodríguez, D., Wijnen, M. and Pakulski, I., World Bank, Washington, US. ISBN: 978-1-4648-0475-5; e-ISBN: 978-1-4648-0476-2 ; <http://dx.doi.org/10.1596/978-1-4648-0475-5>., PDF: <https://openknowledge.worldbank.org/bitstream/handle/10986/22952/9781464804755.pdf>

Guha, S., Hafen, R., Rounds, J., Xia, J., Li, J., Xi, B. and Cleveland, W. S. (2012). Large complex data: divide and recombine (D & R) with RHipe. *Stat*, 1, 53–67.

Haapanen, R., Ek, A. R., Bauer, M. E. and Finley, A. O. (2004). Delineation of forest/nonforest land use classes using nearest neighbor methods. *Remote Sensing of Environment*, 89(3), 265–271.

Han, M. and Liu, B. (2015). Ensemble of extreme learning machine for remote sensing image classification. *Neurocomputing*, 149, 65–70.

Hansen, M. C., DeFries, R. S., Townshend, J. R. G. and Sohlberg, R. (2000). Global land cover classification at 1 km spatial resolution using a classification tree approach. *International journal of remote sensing*, 21(6-7), 1331–1364.

Hansen, M., Dubayah, R. and DeFries, R. (1996). Classification trees: an alternative to traditional land cover classifiers. *International journal of remote sensing*, 17(5), 1075–1081.

Hardin, P. J. (1994). Parametric and nearest-neighbor methods for hybrid classification: a comparison of pixel assignment accuracy. *Photogrammetric Engineering and Remote Sensing*, 60(12), 1439–1448.

Hardin, P. J. and Thomson, C. N. (1992). Fast nearest neighbor classification methods for multispectral imagery. *The Professional Geographer*, 44(2), 191–202.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). The Elements of Statistical Learning Data Mining, Inference and Prediction. New York: Springer, second edition.

Hentze, K., Thonfeld, F. and Menz, G. (2016). Evaluating Crop Area Mapping from MODIS Time-Series as an Assessment Tool for Zimbabwe's "Fast Track Land Reform Programme". *PloS one*, 11(6).

Hogland, J., Billor, N., Anderson, N. (2013) Comparison of standard maximum likelihood classification and polytomous logistic regression used in remote sensing. *European Journal of Remote Sensing* 46, 623-640.

Hooten, M. B. and Wikle, C. K. (2008). A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian Collared-Dove. *Environmental and Ecological Statistics*, 15(1), 59–70.

Hooten, M. B. and Wikle, C. K. (2010). Statistical Agent-Based Models for Discrete Spatio-Temporal Systems. *Journal of the American Statistical Association*, 105(October 2014), 236–248.

Hooten, M. B., Wikle, C. K., Dorazio, R. M. and Royle, J. A. (2007). Hierarchical spatio-temporal matrix models for characterizing invasions. *Biometrics*, 63(2), 558–567.

Hoque, Phinn, S. et al. (2016) Assessing tropical cyclone impacts using object-based moderate spatial resolution image analysis: a case study in Bangladesh. International Journal of Remote Sensing, 37 22: 5320-5343.

Hoque, Phinn, S. et al. (2017): Modelling tropical cyclone risks for present and future climate change scenarios using geospatial techniques. International Journal of Digital Earth, 1-18.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. New York: Springer.

Huang, C., Davis, L. S. and Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4), 725–749.

Huang, G., Huang, G. B., Song, S. and You, K. (2015). Trends in extreme learning machines: A review. *Neural Networks*, 61, 32–48.

Huang, X. and Zhang, L. (2013). An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. *Geoscience and Remote Sensing, IEEE Transactions on*, 51(1), 257–272.

Irwin, M. E., Cressie, N. and Johannesson, G. (2002). Spatial-temporal nonlinear filtering based on hierarchical statistical models. *Test*, 11(2), 249–302.

Walter, V. (2003). Object-based classification of remote sensing data for change detection. *ISPRS Journal of photogrammetry and remote sensing*, 58(3), 225–238.

Jackson, Q., Landgrebe, D. and Others (2002). Adaptive Bayesian contextual classification based on Markov random fields. *Geoscience and Remote Sensing, IEEE Transactions on*, 40(11), 2454–2463.

Jacques, J. and Preda, C. (2014). Functional data clustering: a survey. *Advances in Data Analysis and Classification*, 8(3), 231–255.

Jin, X., Banerjee, S. and Carlin, B. P. (2007). Order-Free Co-Regionalized Areal Data Models with Application to Multiple-Disease Mapping. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 69(5), 817–838.

Jin, X., Carlin, B. P. and Banerjee, S. (2005). Generalized Hierarchical Multivariate CAR Models for Areal Data. *Biometrics*, 61(4), 950–961.

Julea, A., Méger, N., Bolon, P., Rigotti, C., Doin, M.P., Lasserre, C., Trouvé, E. and Lazarescu, V.N. (2011). Unsupervised spatiotemporal mining of satellite image time series using grouped frequent sequential patterns. *IEEE Transactions on Geoscience and Remote Sensing*, 49(4), 1417–1430.

Julea, A., Méger, N., Rigotti, C., Trouvé, E., Jolivet, R. and Bolon, P. (2012). Efficient Spatio-temporal Mining of Satellite Image Time Series for Agricultural Monitoring. *Trans. MLDM*, 5(1), 23–44.

Katzfuss, M. and Cressie, N. (2011). Spatio-Temporal Smoothing and EM Estimation for Massive Remote-Sensing Data Sets. *Journal of Time Series Analysis*, 32(4), 430–446.

Katzfuss, M. and Cressie, N. (2012). Bayesian hierarchical spatio-temporal smoothing for very large datasets. *Environmetrics*, 23(1), 94–107.

Katzfuss, M., Stroud, J. R. and Wikle, C. K. (2016). Understanding the ensemble Kalman filter. *The American Statistician*.

Kavzoglu, T. and Mather, P. M. (2003). The use of backpropagating artificial neural networks in land cover classification. *International Journal of Remote Sensing*, 24(23), 4907–4938.

Kavzoglu, T. and Reis, S. (2008). Performance analysis of maximum likelihood and artificial neural network classifiers for training sets with mixed pixels. *GIScience & Remote Sensing*, 45(3), 330–342.

Kim, H., Sun, D. and Tsutakawa, R. K. (2001). A Bivariate Bayes Method for Improving the Estimates of Mortality Rates with a Twofold Conditional Autoregressive Model. *Journal of the American Statistical Association*, 96(456), 1506–1521.

Kittler, J., Hatef, M., Duin, R. P. W. and Matas, J. (1998). On combining classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(3), 226–239.

Kneip, A. and Sarda, P. (2011). Factor Models and Variable Selection in High-Dimensional Regression Analysis. *The Annals of Statistics*, 39(5), 2410–2447.

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in medicine*, 19(17-18), 2555–2567.

Knorr-Held, L. and Besag, J. (1998). Modelling risk from a disease in time and space. *Statistics in medicine*, 17(18), 2045–2060.

Knorr-Held, L. and Rue, H. (2002). On Block Updating in Markov Random Field Models for Disease Mapping. *Scandinavian Journal of Statistics*, 29(4), 597–614.

Kong, Z. and Cai, Z. (2007). Advances of Research in Fuzzy Integral for Classifiers' fusion. *Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2007. SNPD 2007. Eighth ACIS International Conference on*, volume 2, pp. 809–814.

Kuching, S., Shafri, H. Z. M., Suhaili, A. and Mansor, S. (2007). The Performance of Maximum Likelihood, Spectral Angle Mapper, Neural Network and Decision Tree Classifiers in Hyperspectral Image Analysis. *Journal of Computer Science*, 3(6), 419–423.

Kuhn, H. (1955). The Hungarian Method for the Assignment Problem. *Naval Research Logistics Quarterly*, 2(1-2), 83–97.

Kussul, N., Skakun, S. and Kussul, O. (2006). Comparative analysis of neural networks and statistical approaches to remote sensing image classification. *International Journal of Computing*, 5(2), 93–99.

Lagazio, C., Dreassi, E. and Biggeri, A. (2001). A hierarchical Bayesian model for space time variation of disease risk. *Statistical Modelling*, 1(01), 17–29.

Lam, L. (2000). Classifier combinations: implementations and theoretical issues. In *Multiple classifier systems*, pp. 77–86. Springer.

Lange, T., Mosler, K. and Mozharovskyi, P. (2014). Fast nonparametric classification based on data depth. *Statistical Papers*, 55(1), 49–69.

Latifovic, R., Trishchenko, A. P., Chen J., Park W.B., Khlopenkov, K. V., Fernandes, R., Pouliot, D., Ungureanu, C., Luo, Y., Wang, S., Davidson, A., Cihlar, J., (2005). Generating historical AVHRR 1 km baseline satellite data records over Canada suitable for climate change studies. *Canadian Journal of Remote Sensing*, 31(5), 324–346.

Lawley, E.F., Lewis, M.M., and B. Ostendorf. (2011). Environmental zonation across the Australian arid region based on long-term vegetation dynamics. *Journal of Arid Environments*, 75: 576–585.

Lawrence, R.L. and Wright, A. (2001) Rule-based classification systems using Classification and Regression Tree (CART) analysis. *Photogrammetric Engineering and Remote Sensing*, 67, 1137–1142.

Lawson, A. B. (2013). *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. CRC press, second edition.

Le, N. D. and Zidek, J. V. (2006). *Statistical analysis of environmental space-time processes*. Springer Science & Business Media.

Lee, T.-W. and Lewicki, M. S. (2002). Unsupervised image classification, segmentation, and enhancement using ICA mixture models. *Image Processing, IEEE Transactions on*, 11(3), 270–279.

Lee, T.-W., Lewicki, M. S. and Sejnowski, T. J. (2000). ICA mixture models for unsupervised classification of non-Gaussian classes and automatic context switching in blind signal separation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(10), 1078–1089.

Lee, Y., Lin, Y. and Wahba, G. (2001). Multicategory support vector machines. In *Proceedings of the 33rd Symposium on the Interface*. Costa Mesa, CACiteseer, TECHNICAL REPORT NO. 1043.

Lee, Y., Lin, Y. and Wahba, G. (2004). Multicategory support vector machines: Theory and application to the classification of microarray data and satellite radiance data. *Journal of the American Statistical Association*, 99(465), 67–81.

Lewis, A., Lymburner, L., Purss, M.B.J., Brooke, B., Evans, B., Dekker, A.G., Irons, J.R., Minchin, S., Mueller, N., Oliver, S., Roberts, D., Ryan, B., Thankappan, M., Woodcock, R. & Wyborn (2016), Rapid, high-resolution detection of environmental change over continental scales from satellite data - the Earth Observation Data Cube.

Li, C.H., Kuo, B.C., Lin, C.T. and Huang, C.S. (2012a). A spatial-contextual support vector machine for remotely sensed image classification. *Geoscience and Remote Sensing, IEEE Transactions on*, 50(3), 784–799.

Li, J., Marpu, P. R., Plaza, A., Bioucas-Dias, J. M. and Benediktsson, J. A. (2012b). A New Multiple Classifier System for Semi-supervised Analysis of Hyperspectral Images. In *ICPRAM (1)*, pp. 406–411.

Li, X., Liu, X. and Yu, L. (2014). Aggregative model-based classifier ensemble for improving land use/cover classification of Landsat TM Images. *International Journal of Remote Sensing*, 35(4), 1481–1495.

Liao (2017) The Impact of Energy Consumption on the Surface Urban Heat Island in China's 32 Major Cities. *Remote Sensing* 9, 250.

Lindgren, F. and Rue, H. (2015). Bayesian Spatial Modelling with R-INLA. *Journal of Statistical Software*, 63(19), 1-25.

Lindgren, F., Rue, H. and Lindstrom, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach [with Discussion]. *Journal of the Royal Statistical Society Series B: Methodology*, 73(4), 423–498.

Liu, C., Ray, S., Hooker, G., Friedl, M. (2012) Functional factor analysis for periodic remote sensing data. *The Annals of Applied Statistics* 6(2), 601-624.

Lloyd, C. D. (2010). *Local models for spatial analysis*. CRC Press, second edition.

Lopatin, J., Dolos, K., Hernandez, H.J., Gelleguillos, M., Fasshacht, F.E. (2016). Comparing Generalized Linear Models and random forest to model vascular plant species richness using LiDAR data in a natural forest in central Chile. *Remote Sensing of the Environment* 173, 200-2010.

López-Granados, F., Peña-Barragán, J.M., Jurado-Expósito, M., Francisco- Fernández, M., Cao, R., Alonso-Betanzos, A., and Fontenla-Romero, O. (2008). Multispectral classification of grass weeds and wheat (*Triticum durum*) using linear and nonparametric functional discriminant analysis and neural networks. *Weed Research*, 48, 28–37.

Loew, F., Michel, U., Dech, S. and Conrad, C. (2013). Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using support vector machines. *ISPRS Journal of Photogrammetry and Remote Sensing*, 85, 102–119.

Loew, F., Schorcht, G., Michel, U., Dech, S. and Conrad, C. (2012). Per-field crop classification in irrigated agricultural regions in middle Asia using random forest and support vector machine ensemble. In *SPIE Remote Sensing*, (8538). International Society for Optics and Photonics.

Lu, D. and Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5), 823–870.

Lu, D., Batistella, M., Moran, E. and Mausel, P. (2004). Application of spectral mixture analysis to Amazonian land-use and land-cover classification. *International Journal of Remote Sensing*, 25(23), 5345–5358.

Lu, D., Moran, E. and Batistella, M. (2003). Linear mixture model applied to Amazonian vegetation classification. *Remote sensing of environment*, 87(4), 456–469.

M. K., Plant, R. E. and Six, J. (2011). Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sensing of Environment*, 115(6), 1301–1316.

MacNab, Y. C. and Dean, C. B. (2001). Autoregressive spatial smoothing and temporal spline smoothing for mapping rates. *Biometrics*, 57(3), 949–956.

Marchesi, S. and Bruzzone, L. (2009). ICA and kernel ICA for change detection in multi-spectral remote sensing images. In *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009*, volume 2, pp. II-980. IEEE.

Marley, J., Defina, R., Traeger, K., Elazar, D., Amarasinghe, A., Biggs, G., and Tam, S-M. 2016. Investigative Pilot Report (unpublished).

Martino, S., Aas, K., Lindqvist, O., Neef, L. R. and Rue, H. (2011). Estimating stochastic volatility models using integrated nested Laplace approximations. *The European Journal of Finance*, 17(7), 487–503.

Martins, T. G., Simpson, D., Lindgren, F. and Rue, H. (2013). Bayesian computing with INLA: New features. *Computational Statistics and Data Analysis*, 67, 68–83.

Maselli, F., Rodolfi, A. and Conese, C. (1996). Fuzzy classification of spatially degraded Thematic Mapper data for the estimation of sub-pixel components. *International Journal of Remote Sensing*, 17(3), 537–551.

Mather, P., Tso, B. (2016) Classification Methods for Remotely Sensed Data, Second Edition. CRC Press.

Mathur, A. and Foody, G. M. (2008). Crop classification by support vector machine with intelligently selected training data for an operational application. *International Journal of Remote Sensing*, 29(8), 2227–2240.

McIver, D. K. and Friedl, M. A. (2002). Using prior probabilities in decision-tree classification of remotely sensed data. *Remote Sensing of Environment*, 81(2), 253–261.

McRoberts et al. (2007) Estimating areal means and variances of forest attributes using the k-Nearest Neighbors technique and satellite imagery. *Remote Sensing of the Environment* 111.

Melgani, F. and Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8), 1778–1790.

Mello, M.P., Risso, J., Atzberger, C., Aplin, P., Pebesma, E., Vieira, C.A.O., Rudorff, B.F.T. (2013) Bayesian networks for raster data (BayNeRD): plausible reasoning from observations. *Remote Sensing* 5(11), 5999-6025.

Mertens, K. C., Verbeke, L. P. C., Westra, T. and De Wulf, R. R. (2004). Sub-pixel mapping and sub-pixel sharpening using neural network predicted wavelet coefficients. *Remote Sensing of Environment*, 91(2), 225–236.

Mishra, N. S., Ghosh, S. and Ghosh, A. (2012). Fuzzy clustering algorithms incorporating local information for change detection in remotely sensed images. *Applied Soft Computing*, 12(8), 2683–2692.

Moreno-Seco, F., Inesta, J. M., De León, P. J. P. and Micó, L. (2006). Comparison of classifier fusion methods for classification in pattern recognition tasks. *Structural, Syntactic, and Statistical Pattern Recognition*, pp. 705–713. Springer.

Morisette, J.T., Khorram, S. (1997) An introduction to using generalized linear models to enhance satellite-based change detection. *Geoscience and Remote Sensing. IGARSS '97. Remote Sensing – A Scientific Vision for Sustainable Development.* 1997 IEEE International.

Morisette, J.T., Khorram, S. and Mace, T. (1999) Land-cover change detection enhanced by generalized linear models. *International Journal of Remote Sensing* 20(14), 2703–2721.

Moser, G. and Serpico, S. B. (2013). Combining support vector machines and markov random fields in an integrated framework for contextual image classification. *Geoscience and Remote Sensing, IEEE Transactions on*, 51(5), 2734–2752.

Moser, G., Serpico, S. B. and Benediktsson, J. A. (2013). Land-cover mapping by Markov modeling of spatial–contextual information in very-high-resolution remote sensing images. *Proceedings of the IEEE*, 101(3), 631–651.

Mountrakis, G., Im, J. and Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247–259.

Muff, S., Riebler, A. and Held, L. (2015). Bayesian analysis of measurement error models using integratednestedLaplaceapproximations. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 231–252.

Mugglin, A. S., Cressie, N. and Gemmell, I. (2002). Hierarchical statistical modelling of influenza epidemic dynamics in space and time. *Statistics in medicine*, 21(18), 2703– 21.

Nakamura, R., Osaku, D., Levada, A., Cappabianco, F., Falcao, A. and Papa, J. (2013). OPF-MRF: Optimum-pathforestandmarkovrandomfieldsforcontextual-basedimage classification. In *Computer Analysis of Images and Patterns*, pp. 233–240. Springer.

Neagoe, V.-E., Chirila-Berbentea, V. (2016) Improved Gaussian mixture model with expectation-maximization for clustering of remote sensing imagery. *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*, DOI: [10.1109/IGARSS.2016.7729792](https://doi.org/10.1109/IGARSS.2016.7729792).

Newlands, N.K., Zamar, D.S., Kouadio, L.A., Zhang, Y., Chipanshi, A., Potgieter, A., Toure, S. and Hill, H.S. (2014). An integrated, probabilistic model for improved seasonal forecasting of agricultural crop yield under environmental uncertainty. *Frontiers in Environmental Science*, 2, 17.

Norton, J. D. and Niu, X.-F. (2009). Intrinsically Autoregressive Spatiotemporal Models With Application to Aggregated Birth Outcomes. *Journal of the American Statistical Association*, 104(486), 638–649.

Ohmann, J. L. and Gregory, M. J. (2002). Predictive mapping of forest composition and structure with direct gradient analysis and nearest-neighbor imputation in coastal Oregon, USA. *Canadian Journal of Forest Research*, 32(4), 725–741.

Okujeni, A.O. and van der Linden, S. (2014) A Comparison of Advanced Regression Algorithms for Quantifying Urban Land Cover. *Remote Sensing* 6, 6324–6346.

Otupei, J. R. and Blaschke, T. (2010). Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation*, 12(SUPPL. 1), 27–31.

Pal, M. and Mather, P. M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, 86(4), 554–565.

Pal, M. and Mather, P. M. (2005). Support vector machines for classification in remote sensing. *International Journal of Remote Sensing*, 26(5), 1007–1011.

Paola, J. D. and Schowengerdt, R. A. (1995). A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery. *International Journal of remote sensing*, 16(16), 3033–3058.

Paola, J. D. and Schowengerdt, R. A. (1997). The effect of neural-network structure on a multispectral land-use/land-cover classification. *Photogrammetric Engineering and Remote Sensing*, 63(5), 535–544.

Pedley, M. I. and Curran, P. J. (1991). Per-field classification: an example using SPOT HRV imagery. *Remote Sensing*, 12(11), 2181–2192.

Pena et al. (2016) Object based image classification of summer crops with machine learning methods. *Remote Sensing* 6, 5019.

Pen˜a-Barraga˜n, J. M., Ngugi, M. K., Plant, R. E., Six, J., PeAa-BarragAn, J. M., Ngugi, Penalosa, M. A. and Welch, R. M. (1996). Feature selection for classification of polar regions using a fuzzy expert system. *Remote Sensing of Environment*, 58(1), 81–100.

Petropoulos, G. P., Vadrevu, K. P., Xanthopoulos, G., Karantounias, G. and Scholze, M. (2010). A comparison of spectral angle mapper and artificial neural network classifiers combined with Landsat TM imagery analysis for obtaining burnt area mapping. *Sensors*, 10(3), 1967–1985.

Pettitt, A. N., Weir, I. S. and Hart, A. G. (2002). A conditional autoregressive Gaussian process for irregularly spaced multivariate data with application to modelling large sets of binary data. *Statistics and Computing*, 12(4), 353–367.

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E. and Wulder, M.A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148, 42–57.

Phinn, S.R., Dekker, A.G., Brando, V.E., and Roelfsema, C.M. (2005). Mapping water quality and substrate cover in optically complex coastal and reef waters: an integrated approach, *Marine Pollution Bulletin*, 51, (1–4), 459-469. Retrieved from: <http://www.sciencedirect.com/science/article/pii/S0025326X04003583>

Pitt, Michael, K. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446), 590–599.

Planet. (2016). About Planet. Retrieved from <https://www.planet.com/about/>

Platt, R. V. and Rapoza, L. (2008). An Evaluation of an Object-Oriented Paradigm for Land Use/Land Cover Classification. *The Professional Geographer*, 60(1), 87–100. Pollock, D. (2009). Trends in economic time series. Retrieved from: <http://www.le.ac.uk/users/dsgp1/COURSES/TSERIES/1TRENDS.PDF>

Potgieter, A., Apan, A., Dunn, P., and Hammer, G. (2007). Estimating crop area using seasonal time series of Enhanced Vegetation Index from MODIS satellite imagery. *Crop and Pasture Science*, 58(4), 316-325.

Potgieter, A., Apan, A., Hammer, G. and Dunn, P. (2011). Estimating winter crop area across seasons and regions using time-sequential MODIS imagery. *International journal of remote sensing*, 32(15), 4281-4310.

Potgieter, A.B., Lawson, K. and Huete, A.R. (2013). Determining crop acreage estimates for specific winter crops using shape attributes from sequential MODIS imagery. *International Journal of Applied Earth Observation and Geoinformation*, 23, 254-263.

Preda, C., Saporta, G. and Le've der, C. (2007). PLS classification of functional data. *Computational Statistics*, 22(2), 223–235.

PRIMS. (2016). Retrieved from <http://philippinericeinfo.ph>

Pringle, M., Schmidt, M. and Tindall, D. (2018): Multi season, multi sensor time series modelling based on geostatistical concepts - to predict broad groups of crops. Submitted to Remote Sensing of Environment.

Qiu, F. (2008). Neuro-fuzzy based analysis of hyperspectral imagery. *Photogrammetric Engineering & Remote Sensing*, 74(10), 1235–1247.

Qi, Y. and J. Zhang. (2009). (2D) PCALDA: An efficient approach for face recognition. *Applied Mathematics and Computation*, 213: 1–7.

R Development Core Team (2011). R: A Language and Environment for Statistical Computing. Vienna, Austria: the R Foundation for Statistical Computing. ISBN: 3-900051-07-0. Retrieved from <http://www.R-project.org/>.

Rahman, A. F. R. and Fairhurst, M. C. (2003). Multiple classifier decision combination strategies for character recognition: A review. *Document Analysis and Recognition*, 5(4), 166–194.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. New York: Springer, second edition.

Ranawana, R. and Palade, V. (2006). Multi-classifier systems: Review and a roadmap for developers. *Int. J. Hybrid Intell. Syst.*, 3(1), 35–61.

Rani, S., and Sikka, G. (2012). Recent techniques of clustering of time series data: a survey. *International Journal of Computer Applications* 52(15), 1-9.

Reed, B.C., Brown, J.F., VanderZee, D. et al. (1994) Measuring phenological variability from satellite imagery. *Journal of Vegetation Science* 5(5), Applications of Remote Sensing and Geographic Information Systems in Vegetation Science, pp. 703-714.

Richards, J. (2013) *Remote Sensing Digital Image Analysis*, Springer, Berlin Scannapieco, M., Virgillito, A., and D. Zardetto. 2013. *Placing Big Data in Official Statistics: A Big Challenge?* Accessed: https://ec.europa.eu/eurostat/cros/system/files/NTTS2013fullPaper_214.pdf

Richardson, S., Abellán, J. J. and Best, N. (2006). Bayesian spatio-temporal analysis of joint patterns of male and female lung cancer risks in Yorkshire (UK). *Statistical Methods in Medical Research*, 15(4), 385–407.

Riebler, A., Held, L. and Rue, H. (2012). Estimation and Extrapolation of Time Trends in Registry Data – Borrowing Strength from Related Populations. *The Annals of Applied Statistics*, 6(1), 304–333.

Ripley, B. D. (1977). Modelling Spatial Patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(2), 172–212.

Ripley, B. D. (1988). *Statistical Inference for Spatial Processes*. Cambridge University Press. Ripley, B. D. (1994). Neural networks and related methods for classification. *Journal of the Royal Statistical Society. Series B (Methodological)*, 409–456.

Ripley, B. D. (1996). *Pattern recognition and neural networks*. Cambridge university press.

Roos, M. and Held, L. (2011). Sensitivity analysis in Bayesian generalized linear mixed models for binary data. *Bayesian Analysis*, 6(2), 259–278.

Roy, M., Ghosh, S. and Ghosh, A. (2014). A novel approach for change detection of remotely sensed images using semi-supervised multiple classifier system. *Information Sciences*, 269, 35–47.

Royle, J. A. and Berliner, L. M. (1999). A Hierarchical Approach to Multivariate Spatial Modeling and Prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, 4(1), 29–56.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. CRC Press.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B*, 71(2), 319–392.

Ruiz-Cárdenas, R., Krainski, E. T. and Rue, H. (2012). Direct fitting of dynamic models using integrated nested Laplace approximations - INLA. *Computational Statistics and Data Analysis*, 56(6), 1808–1828.

Sahu, S. K. and Mardia, K. V. (2005). Recent Trends in Modeling Spatio-Temporal Data. Sain, S. R. and Cressie, N. (2007). A spatial model for multivariate lattice data. *Journal of Econometrics*, 140(1), 226–259.

Sain, S. R., Furrer, R. and Cressie, N. (2011). A spatial analysis of multivariate output from regional climate models. *Annals of Applied Statistics*, 5(1), 150–175.

Sakamoto, T., Wardlow, B.D., Gitelson, A.A., Verma, S.B., Suyker, A.E., and Arkebauer, T.J. (2010). A two-step filtering approach for detecting maize and soybean phenology with time-series MODIS data. *Remote Sens. Environ.*, 114, 2146–2159.

Sakamoto, T., Wardlow, B.D., and Gitelson, A.A. (2011) Detecting spatiotemporal changes of corn developmental stages in the US Corn Belt using MODIS WDRVI data. *IEEE Trans. Geosci. Remote Sens.*, 49, 1926–1936.

Samat, A., Du, P., Baig, M. H. A., Chakravarty, S. and Cheng, L. (2014). Ensemble learning with multiple classifiers and polarimetric features for polarized SAR image classification. *Photogrammetric Engineering & Remote Sensing*, 80(3), 239–251.

Scannapieco, M., Virgillito, A., and Zardetto, D. (2013). *Placing Big Data in Official Statistics: A Big Challenge?* Accessed: https://ec.europa.eu/eurostat/cros/system/files/NTTS2013fullPaper_214.pdf

Schmidt, M., Pringle, M., Rakhesh, D., Denham, R. and D. Tindall. 2016. A Framework for Large-Area Mapping of Past and Present Cropping Activity Using Seasonal Landsat Images and Time Series Metrics. Retrieved from <http://www.mdpi.com/2072-4292/8/4/312>.

Schniebel et al. (2017) Using Annual Landsat Time Series for the Detection of Dry Forest Degradation Processes in South-Central Angola. *Remote Sensing* 9, 905.

Schroedle, B. and Held, L. (2011a). A primer on disease mapping and ecological regression using INLA. *Computational Statistics*, 26(2), 241–258.

Schroedle, B. and Held, L. (2011b). Spatio-temporal disease mapping using INLA. *Environmetrics*, 22(6), 725–734.

Schroedle, B., Held, L. and Rue, H. (2012). Assessing the Impact of a Movement Network on the Spatiotemporal Spread of Infectious Diseases. *Biometrics*, 68(3), 736–744.

Sengupta, A. and Cressie, N. (2013). Hierarchical Statistical Modeling of Big Spatial Datasets Using the Exponential Family of Distributions. *Spatial Statistics*, 4, 14–44.

- Sengupta, A., Cressie, N., Frey, R. and Kahn, B. (2013). Statistical modelling of MODIS cloud data using the spatial random effects model. *National Institute for Applied Statistics Research Australia*, 3111–3123.
- Shaddick, G. and Wakefield, J. (2002). Modelling multivariate pollutant data at multiple sites. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 51(3), 351–372.
- Shah, C. A., Arora, M. K. and Varshney, P. K. (2004). Unsupervised classification of hyperspectral data: an ICA mixture model based approach. *International Journal of Remote Sensing*, 25(2), 481–487.
- Shah, C. A., Varshney, P. K. and Arora, M. K. (2007) ICA mixture model algorithm for unsupervised classification of remote sensing imagery. *International Journal of Remote Sensing*, 28(8), 1711–1731.
- Shah, C., Arora, M. K., Robila, S., Varshney, P. K. and Others (2002). ICA mixture model based unsupervised classification of hyperspectral imagery. In *Applied Imagery Pattern Recognition Workshop, 2002. Proceedings. 31st*, pp. 29–35. IEEE.
- Shah, V. P., Younan, N. H., Durbha, S. S. and King, R. L. (2010). Feature identification via a combined ICA–wavelet method for image information mining. *Geoscience and Remote Sensing Letters, IEEE*, 7(1), 18–22.
- Shalan, M. A., Arora, M. K. and Ghosh, S. K. (2003). An evaluation of fuzzy classifications from IRS 1C LISS III imagery: a case study. *International Journal of Remote Sensing*, 24(15), 3179–3186.
- Shao, Y. and Lunetta, R. S. (2012). Comparison of support vector machine, neural net- work, and CART algorithms for the land-cover classification using limited training data points. *ISPRS Journal of Photogrammetry and Remote Sensing*, 70, 78–87.
- Sharma, R., Ghosh, A. and Joshi, P. K. (2013). Decision tree approach for classification of remotely sensed satellite data using open source support. *Journal of Earth System Science*, 122(5), 1237–1247.
- Shekhar, S., Schrater, P. R., Vatsavai, R. R., Wu, W. and Chawla, S. (2002). Spatial contextual classification and prediction models for mining geospatial data. *Multimedia, IEEE Transactions on*, 4(2), 174–188.
- Shumway, R. H. and Stoffer, D. S. (2011). *Time series analysis and its applications: with R examples*. New York: Springer Science & Business Media, third edition.
- Sigrist, F., Kuensch, H. R. and Stahel, W. A. (2015). Stochastic partial differential equation based modelling of large space-time data sets. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 77(1), 3–33.
- Singh, K. K., Nigam, M. J., Pal, K. and Mehrotra, A. (2014). A fuzzy kohonen local information C-means clustering for remote sensing imagery. *IETE Technical Review*, 31(1), 75–81.
- Silipo, R. (2015). Seven Techniques for Data Dimensionality Reduction. Accessed: <https://www.knime.org/blog/seven-techniques-for-data-dimensionality-reduction>
- Sohn, Y. and Rebello, N. S. (2002). Supervised and unsupervised spectral angle classifiers. *Photogrammetric Engineering and Remote Sensing*, 68(12), 1271–1280.
- Sohn, Y., Moran, E. and Gurri, F. (1999). Deforestation in North-Central Yucatan (1985- 1995)- Mapping secondary succession of forest and agricultural land use in Sotuta using the cosine of the angle concept. *Photogrammetric engineering and remote sensing*, 65, 947–958.

Somers, B., Asner, G. P., Tits, L. and Coppin, P. (2011). Endmember variability in spectral mixture analysis: A review. *Remote Sensing of Environment*, 115(7), 1603–1616.

Srivastava, P. K., Han, D., Rico-Ramirez, M. A., Bray, M. and Islam, T. (2012). Selection of classification techniques for land use/land cover change investigation. *Advances in Space Research*, 50(9), 1250–1265.

Stassapoulou, A., Petrou, M., Kittler, J. (1998) Application of a Bayesian network in a GIS based decision making system. *International Journal of Geographic Information Science*, 12, 23-46.

Statistics Canada (2015). Integrated Crop Yield Modelling Using Remote Sensing, Agroclimatic Data and Survey Data. Retrieved from: http://www23.statcan.gc.ca/imdb-bmdi/document/5225_D1_T9_V1-eng.htm

Steele, Pål Roe Sundsøy, Carla Pezzulo, Victor A. Alegana, Tomas J. Bird, Joshua Blumenstock, Johannes Bjelland, Kenth Engø-Monsen, Yves-Alexandre de Montjoye, Asif M. Iqbal, Khandakar N. Hadiuzzaman, Xin Lu, Erik Wetter, Andrew J. Tatem, Linus Bengtsson

Steele, J.E. Sundsoy, P.R., Pezzulo, C. et al. (2017) Mapping poverty using mobile phone and satellite data. *Journal of the Royal Society Interface*. Published online 1 February 2017. DOI: 10.1098/rsif.2016.0690

Strickland, C. M., Simpson, D. P., Turner, I. W., Denham, R. and Mengersen, K. L. (2011). Fast Bayesian analysis of spatial dynamic factor models for multitemporal remotely sensed imagery. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 60(1), 109–124.

Stroud, J. R., Müller, P. and Sanso', B. (2001). Dynamic models for spatiotemporal data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 63, 673–689.

Stroud, J. R., Stein, M. L., Lesht, B. M., Schwab, D. J. and Beletsky, D. (2010). An ensemble Kalman filter and smoother for satellite data assimilation. *Journal of the American Statistical Association*, 105(491), 978–990.

Struijs, P., Braaksma, B. and Daas, P.J.H. (2014). Official statistics and Big Data. Big Data and Society. electronic version last access 13072016 <http://bds.sagepub.com/content/1/1/2053951714538417.article-info>

Suliman, A. and Zhang, Y. (2015). A Review on Back-Propagation Neural Networks in the Application of Remote Sensing Image Classification. *Journal of Earth Science and Engineering*, 5, 52–65.

Sweeney, S., Ruseva, T., Estes, L., Evans, T. (2015) Mapping cropland in smallholder-dominated savannas: integrating remote sensing techniques and probabilistic modelling. *Remote Sensing* 7, 15295-15317.

Szuster, B. W., Chen, Q. and Borger, M. (2011). A comparison of classification techniques to support land cover and land use analysis in tropical coastal zones. *Applied Geography*, 31(2), 525–532.

Tam, S-M. (1987). Analysis of a repeated survey using a dynamic linear model. *International Statistical Review*, 55, 63-73.

Tam, S-M. (2015). A Statistical Framework for Analysing Big Data. *Survey Statistician*, 72, 36-51

Tam, S.-M. and Clarke, F. (2015). Big Data, Statistical Inference and Official Statistics. *International Statistical Review*, 83(3), 436-448.

Tau, J. et al. (2016) A study of a Gaussian mixture model for urban land-cover mapping based on VHR remote sensing imagery. *International Journal of Remote Sensing* 37.

Taubenböck, H., Esch, T., Felbier, A., Wisener, M., Roth, A., Dech, S. (2012) Monitoring urbanization in mega cities from space. *Remote Sensing of the Environment* 117, 162-176.

Taubenböck, H., Esch, T., Felbier, A., Wisener, M., Roth, A., Dech, S. (2012) Monitoring urbanization in mega cities from space. *Remote Sensing of the Environment* 117, 162-176.

Thenkabail, P.S. (2015) Remotely Sensed Data Characterization, Classification, and Accuracies. (Ed). CRC Press.

Therneau, T., Atkinson, B. and Ripley, B. D. (2015). rpart: Recursive Partitioning and Regression Trees.

Tierney, L. and Kadane, J. B. (1986). Accurate Approximations for Posterior Moments and Marginal Densities. *Journal of the American Statistical Association*, 81(393), 82–86.

Tompkins, S., Mustard, J. F., Pieters, C. M. and Forsyth, D. W. (1997). Optimization of endmembers for spectral mixture analysis. *Remote Sensing of Environment*, 59(3), 472–489.

Triepke, F. J., Brewer, C. K., Leavell, D. M. and Novak, S. J. (2008). Mapping forest alliances and associations using fuzzy systems and nearest neighbor classifiers. *Remote Sensing of Environment*, 112(3), 1037–1050.

Tripathy, R., Chaudhary, K.N., Nigam, R., Manjunath, K.R., Chauhan, P., Ray, S.S., Parihar, J.S. (2014). Operational semi-physical spectral-spatial wheat yield model development. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-8, 2014 ISPRS Technical Commission VIII Symposium, 09 – 12 December 2014, Hyderabad, India

Tso, B. and Olsen, R. C. (2005). A contextual classification scheme based on MRF model with improved parameter estimation and multiscale fuzzy line process. *Remote Sensing of Environment*, 97(1), 127–136.

Ugarte, M. D., Goicoa, T. and Militino, A. F. (2010). Spatio-temporal modeling of mortality risks using penalized splines. *Environmetrics*, 21, 270–289.

United Nations Economic Commission for Europe (UNECE), (2013). Generic Statistical Business Process Model – GSBPM - (Version 5.0, December 2013). Retrieved from <http://www.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0>

United Nations Economic Commission for Europe (UNECE). (2015). Forty-sixth session, Report of the Global Working Group on Big data for official statistics, E/CN.3/2015/4. Retrieved from <http://unstats.un.org/unsd/statcom/doc15/2015-4-BigData.pdf>

United States Department of Agriculture (USDA) (2016). May report. Retrieved from <http://www.usda.gov/oce/commodity/wasde/>.

Usman, B. (2013). Satellite Imagery Land Cover Classification using K-Means Clustering Algorithm: Computer Vision for Environmental Information Extraction. *Elixir Journal of Computer Science and Engineering* 18671-18675.

Vapnik, V. (1979). Estimation of Dependencies Based on Data.

Verbeke, L. P. C., Vancoillie, F. M. B. and De Wulf, R. R. (2004). Reusing back-propagation artificial neural networks for land cover classification in tropical savannahs. *International Journal of Remote Sensing*, 25(14), 2747–2771.

Ver Hoef, J.M, Temesgen, H. (2013) A Comparison of the Spatial Linear Model to Nearest Neighbor (k-NN) Methods for Forestry Applications. PLoS ONE 8(3) //doi.org/10.1371/journal.pone.0059129

Wackernagel, H. (2003). *Multivariate geostatistics: an introduction with applications*. Berlin: Springer Science & Business Media, third edition.

Wall, M. M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*, 121(2), 311–324.

Waller, L., Carlin, B. P., Xia, H. and Gelfand, A. E. (1997) Hierarchical Spatio-Temporal Mapping of Disease Rates. *Journal of the American Statistical Association*, 92(438), 607–617.

Walsh, S. (2008) Integration of Hyperion Satellite Data and A Household Social Survey to Characterize the Causes and Consequences of Reforestation Patterns in the Northern Ecuadorian Amazon. *Photogrammetric Engineering and Remote Sensing* 74.

Walter, V. (2004). Object-based classification of remote sensing data for change detection.

Wang, F. (1990). Fuzzy supervised classification of remote sensing images. *Geoscience and Remote Sensing, IEEE Transactions On*, 28(2), 194–201.

Wang, Q., Shi, W., Atkinson, P. M. and Li, Z. (2015). Land Cover Change Detection at Subpixel Resolution With a Hopfield Neural Network. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 8(3), 1339–1352.

Warmerdam, F., Kiselev, A., Morissette, D., Butler, H., Shih, K., Neteler, M., Reimer, S., Amici, A., Villeneuve, S., Byrne, M. and Krawczyk, D. (2016). GDAL - Geospatial Data Abstraction Library.

Wendroth et al. (2003). Predicting yield of barley across a landscape: A state-space modeling approach. *Journal of Hydrology*, 272, 250-263.

Wettle, M., Hatmann, K., Heege, T., and A. Mittal. (2013). Satellite derived bathymetry using physics-based algorithms and multispectral satellite imagery, *Asian Association on Remote Sensing*. Accessed: http://www.a-a-r-s.org/acrs/administrator/components/com_jresearch/files/publications/SC02-0919_Full_Paper_ACRS2013_Wettle-et-al_online.pdf

Whiteside, T. and Ahmad, W. (2005). A comparison of object-oriented and pixel-based classification methods for mapping land cover in northern Australia. In *Proceedings of SSC2005 Spatial intelligence, innovation and praxis: The national biennial Conference of the Spatial Sciences Institute*, 1225–1231.

Whittle, P. (1954). On Stationary Processes in the Plane. *Biometrika*, 41(3/4), 434–449.

Wikle, C. K. (2002) A kernel-based spectral model for non-Gaussian spatio-temporal processes. *Statistical Modelling*, 2(4), 299–314.

Wikle, C. K. (2010) Low rank representations as models for spatial processes. In *Hand- book of Spatial Statistics* (Eds A. E. Gelfand, P. J. Diggle, M. Fuentes and P. Guttorp), 89–106. Boca Raton, FL: Chapman & Hall.

Wozniak, M. M., Gran˜a, M. and Corchado, E. (2014). A survey of multiple classifier systems as hybrid systems. *Information Fusion*, 16(1), 3–17.

Wu, W. and Gao, G. (2012). Remote Sensing Image Classification with Multiple Classifiers Based on Support Vector Machines. In *Computational Intelligence and Design (ISCID), 2012 Fifth International Symposium on*, volume 1, pp. 188–191.

Wyborne, L. (2013). High Performance Computing. AusGeo News March 2013 Issue 109. Accessed <http://www.ga.gov.au/ausgeonews/ausgeonews201303/computing.jsp>.

- Xu, K., Wikle, C. K. and Fox, N. I. (2005). A Kernel-Based Spatio-Temporal Dynamical Model for Nowcasting Weather Radar Reflectivities. *Journal of the American Statistical Association*, 100(472), 1133–1144.
- Xu, L., Krzyzak, A. and Suen, C. Y. (1992). Methods of combining multiple classifiers and their applications to handwriting recognition. *Systems, man and cybernetics, IEEE transactions on*, 22(3), 418–435.
- Yang, C., Everitt, J. H. and Murden, D. (2011). Evaluating high resolution SPOT5 satellite imagery for crop identification. *Computers and Electronics in Agriculture*, 75(2), 347–354.
- Yang, X., Blower, J.D., Bastin, L., Lush, V., Zabala, A., Masó, J., Cornford, D., Díaz, P., Lumsden, J. (2013). An integrated view of data quality in Earth observation. *Phil Trans R Soc A* 371: 20120072.<http://dx.doi.org/10.1098/rsta.2012.0072>.
- Yang, C., Everitt, J. H. and Murden, D. (2011). Evaluating high resolution SPOT 5 satellite imagery for crop identification. *Computers and Electronics in Agriculture*, 75(2), 347–354.
- Yang, J., Zhang, D., Frangi, A. F. and Yang, J.Y. (2004). Two-dimensional PCA: A new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26: 131–137.
- Ye, J., Janardan, R., and Q. Li (2005). Two-dimensional linear discriminant analysis, in L. K. Saul, Y. Weiss & L. Bottou (eds), *Advances in Neural Information Processing Systems 17*, MIT Press, Cambridge, MA, 1569–1576.
- Yeom, J-M., and Kim, H-O. (2015) Comparison of NDVIs from GOCI and MODIS Data towards Improved Assessment of Crop Temporal Dynamics in the Case of Paddy Rice. *Remote sensing* (7), 11326 – 11343.
- Yuan et al. (2015) Continuous Change Detection and Classification Using Hidden Markov Model: A Case Study for Monitoring Urban Encroachment onto Farmland in Beijing. *Remote Sensing* 7, 15318- 15339.
- Yuhas, R. H., Goetz, A. F. H. and Boardman, J. W. (1992). Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm. In *Summaries of the third annual JPL airborne geoscience workshop*, volume 1, pp. 147–149. Pasadena, CA: JPL Publication.
- Zhang, D.Q. and Zhou, Z.H. (2005). 2D PCA: Two-directional two-dimensional PCA for efficient face representation and recognition, *Neurocomputing*, 2005, 69: 224–231.
- Zhang, C. and Xie, Z. (2014). Data fusion and classifier ensemble techniques for vegetation mapping in the coastal Everglades. *Geocarto International*, 29(3), 228–243.
- Zhang, J. and Foody, G. M. (2001). Fully-fuzzy supervised classification of sub-urban land cover from remotely sensed imagery: statistical and artificial neural network approaches. *International journal of remote sensing*, 22(4), 615–628.
- Zhang, J., Wei, F., Sun, P., Pan, Y., Yuan, Z. and Yun, Y. (2015a). A Stratified Temporal Spectral Mixture Analysis Model for Mapping Cropland Distribution through MODIS Time-Series Data. *Journal of Agricultural Science*, 7(8), 95–110.
- Zhang, X. and Chen, C. H. (2002). New independent component analysis method using higher order statistics with application to remote sensing images. *Optical Engineering*, 41(7), 1717–1728.

Zhang, Y., Yang, H. L., Prasad, S., Pasolli, E., Jung, J. and Crawford, M. (2015b). Ensemble multiple kernel active learning for classification of multisource remote sensing data. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 8(2), 845–858.

Zhao, H., Yuen, P.C., and Kwok, J.T. (2006). A novel incremental principal component analysis and its application for face recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 36: 873 – 886.

Zheng, Y. and Zhu, J. (2008). Markov chain Monte Carlo for a Spatial-Temporal Autologistic Regression Model. *Journal of Computational and Graphical Statistics*, 17(1), 123–137.

Zhu, H. and Basir, O. (2005). An adaptive fuzzy evidential nearest neighbor formulation for classifying remote sensing images. *Geoscience and Remote Sensing, IEEE Transactions on*, 43(8), 1874–1889.

Zhu, J., Huang, H. and Wu, J. (2005). Modeling Spatial-Temporal Binary Data Using Markov Random Fields. *Journal of Agricultural, Biological, and Environmental Statistics*, 10(2), 212–225.

Zhu, J., Rasmussen, J. G., Moller, J., Aukema, B. H. and Raffa, K. F. (2008). Spatial- Temporal Modeling of Forest Gaps Generated by Colonization From Below- and Above-Ground Bark Beetle Species. *Journal of the American Statistical Association*, 103(481), 162–177.

Список сокращений

Основные термины

DQF (Data Quality Framework) – Система оценки качества данных

EO (Earth Observations) – наблюдения Земли

NGOs (Non-Governmental organizations) – Неправительственные организации

NSOs (National Statistics Offices) – Национальные статистические службы

Технические термины (Дистанционное зондирование)

ARD (Analysis Ready Data) – Готовые к анализу данные

GEO (Geostationary) satellites – Геостационарные спутники

GSD (Ground Sampling Distance) – Расстояние наземного пробоотбора

LAI (Leaf Area Index) – Индекс площади листьев

LEO (Low Earth polar Orbiting) satellites – Спутники на низкой полярной орбите

LIDAR (Light Detection And Ranging) – Технология обнаружения света и дальности действия

MEO (Medium Earth Orbit) satellites – Спутники на средней околоземной орбите

MIR (Mid Infrared) wavelengths – Волны среднего инфракрасного диапазона

MODIS (MODerate-resolution Imaging Spectroradiometer) – Сканирующий спектрорадиометр среднего разрешения

NIR (Nearby Infrared) wavelengths – Волны ближнего инфракрасного диапазона

OLCI (Ocean Land Colour Imager) – Спектрометр для измерения цвета океана и земли

RMSE (Root Mean Squared Error) – Среднеквадратичная ошибка

Bias – Погрешность

CC (Correlation Coefficient) – Коэффициент корреляции

SAR (Synthetic Aperture Radar) – Радар с синтезированной апертурой

SRAL (Synthetic Aperture Radar Altimeter) – Радиолокационный высотомер с синтезированной апертурой

SLSTR (Sea and Land Surface Temperature Radiometer) – Радиометр температуры поверхности моря и суши

SWE (Snow Water Equivalent) – Эквивалент талой воды

SWIR (Shortwave Infrared) Wavelengths – Коротковолновые инфракрасные волны

TIR (Thermal Infrared) wavelengths - Тепловые инфракрасные волны

VIS (Visible) wavelengths – Волны видимого диапазона

Организации и онлайн-ресурсы

CEOS (Комитет по спутникам наблюдения Земли) <http://ceos.org/>

CRESDA (Ресурсный центр спутниковых данных и приложений, Китай) <http://www.cresda.com/EN/>

CSA (Канадское космическое агентство) <http://www.asc-csa.gc.ca/eng/>

CSIRO (Государственное объединение научных и прикладных исследований) <http://www.csiro.au/>

ESA (Европейское космическое агентство) <http://www.esa.int/ESA>

EU (Европейский союз) http://europa.eu/index_en.htm

GA (Geoscience Australia) <http://www.ga.gov.au/>

GEO (Группа по наблюдению Земли) <https://www.earthobservations.org/index.php>

GEOGLAM (GEO Глобальный мониторинг сельского хозяйства)
<https://www.earthobservations.org/geoglam.php>

GEOSS (Глобальная система наблюдения Земли) <http://www.earthobservations.org/geoss.php>

GFOI (Глобальная инициатива по наблюдению за лесами) <http://www.gfoi.org/>

GWG (Глобальная рабочая группа) по использованию больших данных для целей официальной статистики <http://unstats.un.org/unsd/bigdata/>

INPE (Instituto Nacional de Pesquisas Espaciais, Национальный институт космических исследований Бразилии)

www.inpe.br

JAXA (Японское агентство аэрокосмических исследований) <http://global.jaxa.jp/>

NASA (Национальное управление по аeronавтике и исследованию космического пространства) <https://www.nasa.gov/>

NASA Наблюдение Земли: <https://earthdata.nasa.gov/earth-observation-data>

OSCAR (Инструмент анализа и обзора возможностей систем наблюдений) <https://www.wmo-sat.info/oscar/>

База данных OSCAR: <http://www.wmo-sat.info/oscar/spacecapabilities>

UN GGIM (Комитет экспертов ООН по управлению глобальной геопространственной информацией)

<http://ggim.un.org/>

UN SC (Статистическая комиссия ООН) <http://unstats.un.org/unsd/statcom>

UN SDSN (Сеть ООН для выработки решений в области устойчивого развития) <http://unsdsn.org/>

USGS (Геологическая служба США) <https://www.usgs.gov/>

WMO (Всемирная метеорологическая организация) SDGs (Цели устойчивого развития)

<http://public.wmo.int/en/resources/bulletin/wmo-supporting-2030-agenda-sustainable-development>

NCI (Национальная вычислительная инфраструктура) <http://nci.org.au/>

VPAC (Victorian Partnership for Advanced Computing Ltd)