

mgcvUI: интерактивный интерфейс для пакета mgcv (обобщённые аддитивные модели)

Русский перевод, подготовленный с использованием искусственного интеллекта и
прошедший редакционную проверку; английская версия является основной.

William Bert Craytor

Том 1, Выпуск 1 · Лето 2026

Опубликовано: 1 июля 2026

DOI: [10.67330/xtv1pa25](https://doi.org/10.67330/xtv1pa25)

© 2026 William Bert Craytor. Published by Pacific Vista Net.

Эта статья открытого доступа распространяется на условиях [лицензии Creative Commons Attribution 4.0](#)

Аннотация. mgcvUI — графический интерфейс пользователя для пакета R mgcv, который подгоняет обобщённые аддитивные модели (GAM) с помощью штрафных регрессионных сплайнов с автоматическим выбором степени сглаживания. Он предлагает три режима работы — общее прогнозное моделирование, оценку недвижимости и анализ рыночной зоны — и проводит пользователя через импорт данных, настройку сглаживающих членов, подгонку модели, графики частных эффектов и загружаемые отчёты. В статье описаны требования к формату данных, рабочий процесс моделирования, форматы вывода и полный справочник возможностей mgcvUI.

Ключевые слова: GAM, обобщённые аддитивные модели, сглаживающие члены, штрафные регрессионные сплайны, mgcv, R, графический интерфейс, оценка недвижимости

Publisher's Note (July 2026): Shortly before press, mgcvUI was consolidated into earthUI (version 0.11.0 and later), which now provides the GAM workflow described here alongside the MARS and elastic-net workflows. mgcvUI remains available in its final released form (0.3.0) at <https://github.com/wcraytor/mgcvUI>; future development continues in earthUI. This article applies to mgcvUI as released.

Содержание

1	Введение	4
1.1	История и предыстория	4
1.1.1	Обобщённые аддитивные модели	4
1.1.2	Пакет mgcv	4
1.1.3	Сравнение с MARS (Земля) и эластичной сетью (glmnet)	5
1.1.4	earthUI-mgcvUI Конвейер	6
1.1.5	Рекомендуемый рабочий процесс	6
1.2	Что такое mgcvUI?	6
1.3	Три режима работы	7
1.4	Недвижимость — специальные материалы	7
1.5	Начало работы	8
2	Требования к данным для ввода MLS	9
2.1	Формат файла & Структура	9
2.2	Обязательные столбцы для оценки Режим	9
2.3	Особые правила именования столбцов	10
2.4	Качество данных & Полнота данных	10
2.5	Размещение объекта оценки в строке	11

3	Режим общего назначения	12
3.1	Обзор	12
3.2	Рабочий процесс с использованием боковой панели	12
3.3	Импорт из earthUI (Необязательно)	13
3.4	Вкладки главной панели	13
3.5	Сохранение настроек	14
3.6	Темный режим	14
3.7	Язык и регион & Региональные настройки	14
4	Режим оценки	16
4.1	Обработка строк объекта оценки	16
4.2	Дата оценки & Возраст, с которого разрешена продажа	16
4.3	Специальный столбец Обозначения	16
4.4	Обзор корректировок RCA	16
5	Режим анализа рыночной зоны	17
5.1	Отличия от режима оценки	17
5.2	Когда следует использовать режим «Рынок»	17
6	Отбор переменных	18
6.1	Целевая переменная	18
6.2	Преобразование отклика	18
6.3	Таблица предикторов	18
6.4	Определение типов данных & Переопределения	19
6.5	Специальные типы столбцов См.	19
7	Выбор параметров	21
7.1	Предустановки параметров	21
7.2	Семья	21
7.3	Метод	21
7.4	Gamma (коэффициент сглаживания)	22
7.5	Перекрёстная проверка	22
7.6	Выбрать (Сжатие)	22
7.7	Базовые данные по умолчанию	22
7.8	Значение k по умолчанию (базисная размерность)	23
7.9	Тип тензора	23
7.10	Допустимые взаимодействия	23
7.11	Взаимодействия «фактор сглаживание»	24
7.12	Расширенные параметры	24
8	Подгонка модели	25
8.1	Кнопка «Fit»	25
8.2	Что происходит во время подгонки	25
8.3	Общая информация Подгонка & Ход работы	25
9	Вкладки «Результаты»	26
9.1	Данные	26
9.2	Уравнение	26
9.3	Резюме	26
9.4	Важность переменных	27
9.5	Графики вкладов	27
9.5.1	Univariate Smooth Terms — $s(x)$	27

9.5.2	Фактор-сглаживание Термины — $s(x, by = factor)$	27
9.5.3	Interaction Terms — $ti(x_1, x_2)$ and $te(x_1, x_2)$	27
9.5.4	Параметрические члены — линейные и факторные	27
9.6	Корреляция	28
9.7	Диагностика	28
9.8	Корректировки RCA	28
9.9	Аноа	28
9.10	Вывод Mgcsv:	28
9.11	Проверка знаков	28
9.12	Concurvity	29
9.13	Проверка исходных данных	29
10	Загрузка данных	30
10.1	Столбцы вывода	30
10.2	Результаты CQA: балл	30
11	Расчеты RCA& Скачивание	31
11.1	Открытие диалогового окна RCA	31
11.2	Интерполяция баллов CQA	31
11.3	Столбцы вывода	31
12	Таблица сравнения продаж	32
12.1	Диалоговое окно выбора аналогов	32
12.2	Макет книги	32
12.3	Защита листа & Остатки ячейки	32
13	Скачивание отчетов	33
13.1	Шаг 1: Создание отчёта Quarto Отчёт «	33
13.2	Шаг 2: Конвертация Quarto Отчёт «	33
13.3	Содержание отчёта	33
14	Экспорт функций модели	35
14.1	Функции R	35
14.2	Функции Python	35
14.3	Заголовочный файл C++	35
14.4	База данных SQLite	35
15	Сравнение с earthUI и glmnetUI	37
15.1	Основные различия	37
15.2	Общие функции	37
15.3	Рекомендуемый рабочий процесс	38
16	Демо-набор данных: Оценка _ 1.csv	39
16.1	Описание	39
16.2	Столбцы	39
16.3	Рекомендуемое руководство по быстрому запуску	40
17	Системные требования& Устранение неполадок	41
17.1	Поддерживаемые платформы	41
17.2	Версия R:	41
17.3	Необходимые пакеты R:	41
17.4	Дополнительные зависимости	42
17.5	Примечания по конкретным платформам	42

17.5.1 macOS	42
17.5.2 Windows	42
17.5.3 Linux	42
17.6 Плавная деградация	43
17.7 Устранение неполадок в программе «	43
17.7.1 «Параметр PDF недоступен»	44
17.7.2 «Настройки не сохранятся»	44
17.7.3 «График диагностики недоступен»	44
17.7.4 “Отчёт об ошибке: количество элементов, которые необходимо заменить ...”	44
17.7.5 Порт 7880 уже занят	44
17.7.6 Ошибка установки пакета в Linux	44

1 Введение

1.1 История и предыстория

1.1.1 Обобщенные аддитивные модели

Обобщённые аддитивные модели (GAM) были формализованы Тревором Хэсти и Робертом Тибширани в их монографии 1990 года. GAM расширяют линейную модель путём замены каждого линейного члена $\beta_j x_j$ на функцию сглаживания $f_j(x_j)$, в результате чего модель принимает следующий вид:

$$y = \beta_0 + f_1(x_1) + f_2(x_2) + \dots + f_p(x_p) + \varepsilon$$

Каждая функция f_j оценивается на основе данных с помощью регрессионного сплайна с штрафным термином. Процесс сглаживания позволяет модели отражать нелинейные зависимости без необходимости предварительного указания пользователем формы функции. Штрафной термин регулирует степень «волнистости» каждой сглаженной функции, предотвращая переобучение и одновременно обеспечивая достаточную гибкость для отслеживания реальных закономерностей.

1.1.2 Пакет mgcv

Пакет R «mgcv» (Mixed GAM Computation Vehicle) был разработан Саймоном Вудом из Университета Бата (Wood 2017). Впервые выпущенный в 2001 году, он в настоящее время является стандартной реализацией GAM в R и поставляется с каждой установкой R. Лежащие в его основе методы описаны в серии статей, посвящённых стабильной оценке нескольких параметров сглаживания (Wood 2004), сплайнам регрессии «тонкой пластины» (Wood 2003), быстрой стабильной оценке REML (Wood 2011), а также общему выбору сглаженных моделей (Wood, Pya и Säfken 2016). Ключевые особенности включают:

- Автоматический выбор степени сглаживания : параметр сглаживания для каждого члена оценивается в ходе подгонки модели с использованием методов REML (ограниченное максимальное правдоподобие), GCV (обобщённая перекрёстная проверка) или ML (максимальное правдоподобие).
- Несколько типов базисов : сплайны регрессии тонкой пластины (tp), кубические регрессионные сплайны (cr), P-сплайны (ps) и B-сплайны (bs), каждый из которых обладает своими свойствами.
- Сглаживание с помощью тензорного произведения : $te()$ и $ti()$ позволяют обеспечить плавное взаимодействие между переменными без «проклятия размерности».

- Отбор переменных : опция « `select = TRUE` » добавляет дополнительный штраф, который может свести сглаживающие члены к нулю, эффективно осуществляя отбор переменных.
- Диагностика конкурвности : измеряет аналог мультиколлинеарности для сглаживающих членов.

1.1.3 Сравнение с MARS (Земля) и эластичной сетью (glmnet)

Приложение «Earth» и его два модуля постобработки (glmnet и mgcv) используют разные механизмы моделирования. В приведённой ниже таблице приведены основные различия:

Функции	mgcv (GAM)	earth (MARS)	glmnet (эластичная сеть)
Тип модели	Сглаживаемая нелинейная (штрафные сплайны)	Кусочно-линейная (адаптивные сплайны)	Линейная (с регуляризацией)
Нелинейность	Автоматически с помощью сглаживаемых функций (тонкая пластина, кубическая и т. д.)	Автоматически с помощью функций-шарниров с узлами, определяемыми данными	Только через взаимодействия или разложение в базис
Отбор переменных	Опционально с помощью сжатия с двойным штрафом (<code>select=TRUE</code>)	Встроенно с помощью пошагового метода «вперед/назад» с обрезкой по GCV	Встроенно с помощью штрафа « L_1 » (Lasso)
Взаимодействия	Сглаживание тензорным произведением (<code>te()</code> , <code>ti()</code>)	Произведения на основе шарниров (интерпретируемые, до 3-й степени)	Парные векторные произведения (сложно интерпретировать)
Интерпретируемость	Сглаженные кривые частичных эффектов; очень интуитивно понятно	g-функции показывают кусочно-линейные частичные эффекты по каждой переменной	Коэффициенты представляют собой линейные веса; их легко объяснить по отдельности
Лучше всего подходит для	Сглаженные нелинейные зависимости, гибкое моделирование	Автоматическое обнаружение нелинейности и взаимодействий	Высокомерные данные, отбор переменных, небольшие выборки
Корректировки RCA	Сглаживание с помощью частичных эффектов	Кусочно-линейные корректировки с помощью g-функций	Линейные корректировки по каждой переменной
Контроль переобучения	Оценка параметров сглаживания методами REML/GCV	Обрезка на основе GCV	Лямбда, проверенная с помощью перекрёстной проверки

Метод MARS (Multivariate Adaptive Regression Splines) был предложен Джеромом Фридманом в 1991 году. Он позволяет строить кусочно-линейные модели путем адаптивного выбора функций-шарниров и положений их узлов на основе данных. Реализация этого метода в языке R представлена в пакете earth, разработанном Стивеном Милборроу.

Регрессия с эластичной сетью, реализованная в пакете «glmnet» Фридмана, Хастии и Тибширани (2010), сочетает в себе штрафные функции Lasso (L_1) и Ridge (L_2) для одновременного отбора переменных и сжатия коэффициентов.

1.1.4 earthUI–mgcvUI Конвейер

Эффективный рабочий процесс сочетает в себе автоматический поиск узлов в earthUI с сглаживанием оценки в mgcvUI. Когда earthUI экспортирует файл результатов .rds, а mgcvUI импортирует его:

1. Earth обнаруживает узлы — MARS находит основанные на данных положения шарниров (точки изменения) для каждой переменной.
2. mgcvUI использует их в качестве якорей — Узлы earth становятся основой для кубических регрессионных сплайнов (основа cr) в GAM, давая сглаживаемым членам преимущество благодаря положениям узлов, которые отражают реальную структуру данных.
3. mgcv уточняет подгонку — Оценка сплайна с штрафом проводит сглаживание кусочно-линейной подгонки earth, создавая гладкие кривые частичных эффектов, сохраняя при этом общую форму, обнаруженную earth.

Данный процесс объединяет эксплоративное моделирование (earth) с подтверждающим моделированием (GAM), сочетая в себе лучшие черты обоих подходов.

1.1.5 Рекомендуемый рабочий процесс

Для оценки недвижимости и аналогичных задач, требующих понятных и обоснованных моделей:

1. Начните с earthUI, чтобы выявить важные переменные, нелинейные зависимости и взаимодействия в ваших данных. Автоматический выбор базиса и g-функции earthUI позволяют получить отличные первоначальные выводы.
2. Уточните результаты с помощью mgcvUI, если вам нужны более плавные частичные эффекты. Импортируйте узлы из earthUI в mgcvUI для плавного перехода от кусочно-линейных моделей к моделям сглаживания.
3. Используйте glmnetUI, если вам требуется агрессивный отбор переменных (Lasso), когда количество предикторов приближается к числу наблюдений или превышает его, либо если вам нужна чисто линейная модель с регуляризацией для обоснованности результатов.

1.2 Что такое mgcvUI?

mgcvUI — это графический интерфейс пользователя для пакета R «mgcv». Он работает как локальное приложение Shiny — без входа в систему, без сервера и без учетных записей. Вы запускаете его из R, импортируете набор данных (CSV или Excel), настраиваете модель и в интерактивном режиме выполняете подгонку GAM.

Начиная с текущей версии, mgcvUI больше не является отдельно устанавливаемым пакетом: он поставляется в составе пакета earthUI в качестве модуля постобработки GAM, запускаемого командой `earthUI::launch_mgcv()`. Само приложение — а также всё, что описано в данной статье, — осталось без изменений; отличаются лишь установка и команда запуска. Три процедуры (earth, glmnet, mgcv) используют одно дерево проекта, одну базу данных настроек и одну копию вспомогательной инфраструктуры.

Приложение обеспечивает полный рабочий процесс: импорт данных, настройку переменных с возможностью сглаживания членов, подгонку модели с обработкой в фоновом

режиме, графики для диагностики, кривые сглаженных частичных эффектов, а также отчёты, доступные для скачивания в форматах Word, PDF или HTML.

Обобщённые аддитивные модели заменяют линейную зависимость βx на функцию сглаживания $f(x)$ для каждого предиктора. Это означает:

- Нет необходимости задавать функциональные формы — данные сами определяют форму каждой зависимости.
- Сглаживание частичных эффектов — вклад каждого предиктора представляет собой сглаженную кривую, а не прямую линию, что позволяет легко увидеть, как переменная влияет на зависимую переменную при различных значениях.
- Автоматический выбор степени сглаживания — `mgcv` определяет, насколько извилистой должна быть каждая кривая, с помощью REML или GCV, обеспечивая баланс между подгонкой и переобучением.

1.3 Три режима работы

У каждого проекта есть цель, которая выбирается при его создании в разделе 1 боковой панели («Проект») и впоследствии фиксируется. Цель активного проекта определяет, какие инструменты и элементы интерфейса отображаются; чтобы работать в другом режиме, создайте или откройте проект с соответствующей целью. Существуют три режима:

- Общие — Регрессия GAM для любого типа совокупности или набора данных. Это режим по умолчанию. Он обеспечивает полный рабочий процесс GAM без каких-либо дополнений, специфичных для конкретной области.
- Для оценки — Регрессия GAM, адаптированная для оценки недвижимости. Добавляет функции, специфичные для оценки отдельных объектов недвижимости, включая обработку объекта оценки, специальные обозначения столбцов и Метод остаточных ограничений (RCA).
- Анализ рыночной зоны — Регрессия GAM, адаптированная для исследований рыночных зон. Добавляет функции для анализа групп объектов недвижимости в определённом рынке с опциональной обработкой строк, относящихся к конкретному объекту.

Во всех трёх режимах основной механизм моделирования одинаков — вы всегда выполняете подгонку модели GAM с помощью функции «`mgcv::gam()`». Настройка «Purpose» определяет, какие дополнительные инструменты и элементы интерфейса будут доступны.

1.4 Недвижимость — специальные материалы

При выборе опции «Для оценки» или «Анализ рыночной зоны» программа `mgcvUI` активирует ряд функций, предназначенных для анализа рынка недвижимости:

- Обозначения специальных столбцов — Каждому предиктору может быть присвоена специальная роль, такая как «`contract_date`», «`dom`», «`concessions`», «`latitude`», «`longitude`», «`living_area`», «`lot_size`», «`actual_age`», «`effective_age`», «`area`», «`site_dimensions`» или «`display_only`». Эти обозначения определяют, как столбец будет обрабатываться во время подгонки модели и при выводе результатов.
- Столбец «Age» (Срок с момента продажи) — Если столбец обозначен как «`contract_date`» (Срок с момента продажи) и указана «дата оценки» (Дата вступления в силу), `mgcvUI` вычисляет столбец «`sale_age`» (Срок с момента продажи) (количество дней между датой продажи и датой оценки) и использует его в качестве предиктора.
- Вычисления RCA (только при оценке) — В режиме оценки, после подгонки модели `mgcvUI` может вычислить вывод по методу остаточных ограничений (RCA), который даёт корректировки по каждому аналогу, сводные данные по чистым/валовым кор-

ректировкам, а также скорректированную цену продажи недвижимости, подлежащей оценке.

Совет: Режим «Общее назначение» подходит для любых наборов данных — финансовых, научных, инженерных или связанных с недвижимостью. Режимы «Оценка» и «Рынок» просто предоставляют дополнительные удобные функции для специалистов в сфере недвижимости.

1.5 Начало работы

Чтобы воспользоваться mgcvUI:

- 1 . Установка earthUI в R: `install.packages("earthUI")` — процедура mgcv уже включена. (См. раздел «Установка earthUI и vProlog» в статье об earthUI в этом выпуске, где описаны версии для разработки версии и дополнительных компонентов.)
- 2 . Запустите приложение : выполните команду `earthUI::launch_mgcv()` в R. Приложение откроется в вашем веб-браузере на порту 7880.
- 3 . Создайте или откройте проект в разделе 1 боковой панели. Проект определяет цель (Общий, Для оценки или Анализ рыночной зоны) и место проведения работ (страна, штат, округ, город), а также определяет папки ввода и вывода. Проекты хранятся в общем корневом каталоге regProj и доступны для других приложений earthUI и glmnetUI. Открытие проекта с другой целью позволяет переключаться между режимами.
- 4 . Импортируйте данные из раздела 2, выбрав файл CSV или Excel из папки вводных данных активного проекта. Добавьте файлы в эту папку, затем нажмите «Обновить», чтобы отобразить их список.
- 5 . Импорт из earthUI (необязательно) — Раздел 3 позволяет импортировать файл результатов earthUI .rds для определения начальных значений в GAM с помощью узлов, обнаруженных earthUI. Пропустите этот шаг, если у вас нет результатов earthUI .
- 6 . Настройка переменных — выберите целевую переменную и предикторы, установите типы данных, назначьте специальные роли столбцам и пометьте переменные как факторные или линейные.
- 7 . Настройка параметров mgcv — семейство, метод, гамма, тип базиса, k, тип тензора, взаимодействия и другие параметры.
- 8 . Постройте модель — нажмите «Fit Mgcv GAM Model» и просмотрите результаты на главной панели.
- 9 . Экспорт — скачайте прогнозы в формате Excel, сгенерируйте и конвертируйте отчёт Quarto или (в режиме оценки) рассчитайте корректировки RCA и таблицу продаж.

Настройки автоматически сохраняются в базе данных SQLite и восстанавливаются при повторной загрузке того же файла ввода.

2 Требования к данным для ввода MLS

В рамках рабочих процессов по оценке недвижимости и анализу рынка вводные данные обычно получают из экспорта из системы Multiple Listing Service (MLS). В этой главе описывается ожидаемая структура файла и столбцы, которые может использовать mgcvUI.

2.1 Формат файла & Структура

mgcvUI поддерживает файлы в форматах CSV и Excel (.xlsx, .xls). При импорте названия столбцов автоматически преобразуются в формат snake_case — например, «Living SqFt» превращается в «living_sqft», «Contract Date» — в «contract_date», а «Цена продажи» — в «sale_price». Такая нормализация обеспечивает единообразные ссылки на столбцы на протяжении всего рабочего процесса. Разделитель CSV и знак десятичной запятой, используемые при импорте, определяются настройками локали (см. главу 3, «& ные региональные настройки»).

Ваш файл с данными должен представлять собой плоскую таблицу, в которой каждая строка соответствует одному объекту недвижимости, а каждый столбец — одной характеристике. Первая строка файла должна содержать заголовки столбцов.

2.2 Обязательные столбцы для оценки Режим

Хотя mgcvUI работает с любым набором столбцов, для полноценного рабочего процесса оценки рекомендуется использовать следующие столбцы:

Столбец	Специальный тип	Назначение
Цена продажи	(целевая)	Переменная отклика для модели
Дата заключения договора	contract_date	Используется для расчета sale_age (количество дней с даты оценки)
Дата размещения объявления	listing_date	Используется вместе с датой заключения договора для расчета DOM, если столбец DOM отсутствует
Дни на рынке	dom	Дни на рынке; отображается в экспортируемых данных
Уступки	concessions	Уступки при продаже; чистая цена продажи (Net SP) = цена продажи – Уступки
Жилая площадь (SF)	living_area	Позволяет рассчитывать остатки на квадратный фут (residual_sf, cqa_sf)
Размер участка	lot_size	Столбец «Размер участка»
Размеры участка	site_dimensions	Группируется по размеру участка
Широта	latitude	Округлено до 3 знаков после запятой
Долгота	longitude	Округлено до 3 знаков после запятой
Идентификатор района	area	Идентификатор рыночного района / микрорайона
Фактический возраст	actual_age	Возраст объекта недвижимости
Эффективный возраст	effective_age	Эффективный возраст объекта недвижимости
Адрес	display_only	Отображается в экспортируемых данных; исключается из модели

Названия столбцов в таблице могут быть на иностранном языке — «специальные» названия должны быть на английском языке, чтобы программа R могла обрабатывать их особым образом. В противном случае указанные названия столбцов будут отображаться в регрессионных моделях, графиках и, при проведении оценок, в выводах.

Не все столбцы являются обязательными. mgcvUI адаптируется — если какой-либо столбец отсутствует, соответствующий объект просто не учитывается. Однако для моделей ценообразования на рынке недвижимости настоятельно рекомендуется использовать определенные столбцы:

- 1 . Срок продажи — количество дней между датой заключения договора купли-продажи и датой оценки. Если используется многолетняя история продаж, `sale_age` часто играет ключевую роль.
- 2 . Жилая площадь — также известна под такими названиями, как «Living Sqft», «GLA» (общая жилая площадь).
- 3 . Общее количество ванных комнат — общее количество полных, четвертных, половинных и 3/4 ванных комнат.
- 4 . Гаражные места или площадь гаража — количество гаражных мест или площадь в квадратных футах.
- 5 . Площадь участка — площадь недвижимости.
- 6 . Долгота, широта и , при наличии , идентификатор района для определения географических различий в ценах.

2.3 Особые правила именования столбец

mgcvUI идентифицирует столбцы по их специальным обозначениям типов, а не по названиям столбцов. При экспорте в формат MLS вы можете называть столбцы как угодно — главное, чтобы в таблице «Настройка переменных» (глава 6) был назначен правильный специальный тип.

2.4 Качество данных& Полнота данных

- Отсутствующие значения (NA) : Строки, содержащие значения NA в любом столбце предиктора или целевой переменной, автоматически удаляются перед подгонкой модели. mgcvUI формирует подмножество данных, включающее только столбцы, используемые в модели, поэтому значения NA в неиспользуемых столбцах не приводят к удалению строк.
- Числовые столбцы : Цена продажи, жилая площадь, площадь участка и подобные поля должны быть числовыми.
- Факторные столбцы : Текстовые, TRUE/FALSE и R-факторные столбцы автоматически рассматриваются как категориальные — они могут входить в модель только в качестве факторов. Числовой столбец (например, числовое значение площади или код стиля) становится фактором только при установке флажка «Фактор» ; mgcvUI никогда не выводит категориальность из диапазона значений числового столбца, поэтому дискретные числовые предикторы, такие как `bath_count` , остаются непрерывными (сглаженными). Роль «Special» не зависит от решения о факторах .

Совет: После импорта проверьте столбец «NA» в таблице предикторов. Столбцы с большим количеством пропущенных значений могут привести к исключению строк. Столбцы, в которых количество пропущенных значений превышает $\geq 50\%$, помечаются красным цветом и автоматически исключаются.

2.5 Размещение объекта оценки в строке

В режиме оценки в первой строке должна быть указана недвижимость, которая подлежит оценке. Все остальные строки представляют собой аналоги сделок. Строка с объектом оценки исключается из процесса подгонки модели. После подгонки модели она по-прежнему генерирует прогнозы для этой строки.

В режиме «Анализ рыночной зоны» размещение объекта оценки в строке 1 является необязательным.

В режиме «Общий» специальной обработки строк не предусмотрено — все строки обрабатываются одинаково.

3 Режим общего назначения

3.1 Обзор

Режим «General Purpose» (Общего назначения) является режимом по умолчанию при запуске mgcvUI. Он обеспечивает полный рабочий процесс GAM для любых наборов данных — не только для данных по недвижимости. Вы можете использовать mgcvUI для работы с научными данными, финансового анализа, инженерных исследований или решения любых задач регрессионного анализа, в которых предполагается сглаживание нелинейных зависимостей.

В общем режиме интерфейс не отображает функции, специфичные для рынка недвижимости (специальные столбцы, срок продажи, RCA). Боковая панель упрощена, чтобы сосредоточить внимание на отборе переменных, настройке параметров, подгонке модели и экспорте.

3.2 Рабочий процесс с использованием боковой панели

Боковая панель состоит из пронумерованных сворачиваемых разделов, которые помогают вам пройти весь путь от выбора проекта до экспорта отчёта. Большинство разделов отображаются только при открытии активного проекта, а некоторые из них зависят от цели проекта. Папка для вывода данных определяется на основе активного проекта — отдельного поля для указания папки вывода данных нет.

1. Проект — создание или открытие проекта. В проекте задаются цель и место проведения работ (страна, штат, округ, город), а также папка ввода (`<os>_in/`) и папка вывода. Проекты хранятся в общем корневом каталоге `regProj` (по умолчанию `~/regProj`, можно переопределить с помощью переменной среды `REGPROJ_ROOT` или раскрывающегося списка «Настройки») и совместно используются с родственными приложениями `earthUI` и `glmnetUI`, поэтому модели из всех трёх приложений хранятся в одном месте. Нажмите «Закрыть проект», чтобы переключиться.

2. Импорт данных — Выберите файл CSV или Excel из папки ввода текущего проекта. Добавьте файлы в папку с помощью Finder/Проводника, затем нажмите «Обновить», чтобы отобразить их список. Для файлов Excel с несколькими листами появится окно выбора листа. Названия столбцов автоматически преобразуются в формат «snake _».

3. Импорт из `earthUI` (необязательно) — Импортируйте файл результатов `earthUI .rds`, чтобы установить начальные значения для GAM с помощью координат узлов, обнаруженных `earthUI` (а также для отбора переменных). Этот шаг является необязательным — пропустите его, если у вас нет результатов `earthUI` для импорта.

4. Настройка переменных — выбор целевой переменной, преобразование отклика (без преобразования, логарифмическое, логарифмическое с основанием 10), таблица предикторов с флажками «Включить», «Фактор» и «Линейный». Столбец «Специальное» отображается только в режимах «Оценка» и «Рынок». Подробную информацию см. в главе 6.

5. Параметры вызова `Mgcv` — Все настройки модели: предустановки параметров, семейство, метод, гамма, перекрёстная проверка, выбор, тип базиса, k , тип тензора, матрица взаимодействия и расширенные параметры. Полное справочное руководство по параметрам см. в главе 7.

6. Подгонка модели `Mgcv GAM` — кнопка, запускающая модель. Результаты отображаются на вкладках справа.

7. Скачать результаты — экспортирует прогнозы, остатки, балы CQA и вклады по каждой переменной в виде файла Excel в папку вывода проекта. Доступно во всех режимах после подгонки модели.

8. Расчет корректировок RCA & Скачать — Вычисляет корректировки RCA для каждой переменной по каждому аналогу и интерполирует остаток по объекту оценки с помощью CQA. Доступно только в режимах «Оценка» и «Рынок», после подгонки модели.

9. Создать таблицу продаж & Скачать — Создает таблицу сравнения продаж, предлагая аналогичные объекты с учетом процента валовой корректировки. Отображается только в режиме оценки.

10. Создать отчет Quarto — создает автономный набор отчета Quarto (.qmd) — со всеми графиками и таблицами — в папке вывода проекта. Функция становится доступной после подгонки модели.

11. Конвертировать отчет Quarto — преобразует сгенерированный набор .qmd в формат HTML, Word или PDF. Для создания PDF-файла требуется установка LaTeX; если она не обнаружена, опция PDF будет скрыта. См. главу 13.

В других режимах шаги, предназначенные для конкретных задач, опущены (а оставшиеся шаги перенумерованы): в режиме «Рынок» отсутствует шаг «Таблица продаж», а в режиме «Общий» отсутствуют оба шага — «RCA» и «Таблица продаж».

3.3 Импорт из earthUI (Необязательно)

Раздел 3 боковой панели — «Импорт из earthUI» — позволяет импортировать файл результатов earthUI .rds. Это активирует конвейер earth-mgcv: позиции узлов, определяемые данными earth, становятся опорными точками для членов процесса сглаживания GAM.

При импорте результата из earthUI mgcvUI:

- Отображает сводку, содержащую целевую переменную, коэффициент R-квадрат, количество членов и положения узлов для каждой переменной.
- Автоматически выбирает предустановку параметров «Earth Pipeline» (кубическая регрессионная сплайн-основа, гамма = 1,4).
- Устанавливает целевую переменную в соответствии с целевой переменной «earth».
- Предварительно устанавливает флажок «Include» для предикторных переменных earth.
- Предварительно устанавливает флажки «Factor» для категориальных переменных earth и «Linear» для линейных предикторов earth.
- Предварительно заполняет взаимодействия, обнаруженные earth (выделены желтым цветом в матрице взаимодействий, заблокированы для ручного изменения).
- Предоставляет кнопку «Экспорт узлов в CSV» для загрузки узлов earth в виде файла CSV.

Затем в таблице «Конфигурация переменных» отображается примечание, объясняющее, что отбор переменных («Включить», «Линейная», «Фактор») и узлы сплайна имеют начальное значение из модели Земли. В отличие от glmnetUI, где подгонка glmnet привязана к базису earth, mgcvUI проводит подгонку GAM на основе вашего текущего выбора — earth служит лишь отправной точкой, и все параметры остаются доступными для редактирования.

Кнопка «Очистить» удаляет импортированные данные о заземлении и возвращает устройство в автономный режим.

3.4 Вкладки главной панели

После импорта данных на главной панели отображаются следующие вкладки (вкладки, зависящие от модели, появляются после подгонки):

Вкладка «	» Содержание
Данные	Интерактивная таблица DataTable с импортированными данными. В режимах «Оценка» и «Рынок» разбита на таблицы «Объект оценки» и «Аналоги».
Уравнение	Полная формула GAM, отображенная с помощью MathJax, а также таблица определений функций сглаживания.
Резюме	Ключевые показатели (R^2 , $CV R^2$, Dev. Explained, AIC, n, Кривая сглаживания (сглаженная) и таблицы сглаженных/параметрических термов.
Важность переменных	Горизонтальная гистограмма F-статистики (сглаживание) и значений $ t $ (параметрическая), а также таблица с ранжированием.
Вклад	Интерактивные кривые частичных эффектов с доверительными лентами и маркерами EarthKnot, построенные с помощью Plotly.
Корреляция	Тепловая карта корреляций числовых предикторов (доступна до подгонки модели).
Диагностика	Остатки по сравнению с моделью, Q-Q-диаграмма, гистограмма, зависимость ответа от модели (через gratia). Разброс фактических и прогнозируемых значений.
Корректировки RCA	Гистограммы анализа RCA (режим оценки/рынка, после вычислений).
ANOVA	Таблица ANOVA с объединенными параметрическими и членами сглаживания.
Вывод Mgcsv	Исходный текст: временные параметры, формула, сводка модели, семейство, конкурвитность.
Проверка знаков	Проверка согласованности направлений Earth и GAM (при импорте данных из Earth).
Конкурвитность	Матрицы общей и парной конкурвитности с цветовой кодировкой.
Проверка базиса	gam.check() тесты адекватности базисных измерений.

3.5 Сохранение настроек

mgcsvUI автоматически сохраняет вашу конфигурацию в базе данных SQLite, с ключом, соответствующим имени файла ввода. При повторной загрузке того же файла все настройки восстанавливаются: выбор целевой переменной, флажки предикторов, типы данных, параметры mgcsv и матрица взаимодействия. Цель является свойством активного проекта, а не глобальной настройкой, и последний использованный файл ввода запоминается для каждого проекта и автоматически загружается при повторном открытии проекта.

3.6 Темный режим

Нажмите на значок в виде луны/солнца в правом верхнем углу, чтобы переключаться между темами «Nord Light» и «Nord Dark». Настройки темы сохраняются в localStorage и сохраняются между сессиями. Все элементы пользовательского интерфейса адаптируются к выбранной теме.

3.7 Язык и регион & Региональные настройки

mgcsvUI поддерживает международные стандарты формата чисел и CSV-файлов благодаря системе локалей, основанной на странах. В раскрывающемся списке «Настройки» в заголовке

окна доступны переключатели «Страна» и «Формат бумаги» для более 30 поддерживаемых стран. Каждая предустановка настраивает:

- Разделитель в файлах CSV — запятая (,) для США/Великобритании/Японии или точка с запятой (;) для большинства стран Европы.
- Десятичный знак — точка (.) или запятая (,).
- Разделитель тысяч — запятая (США/Великобритания/Япония), точка (Германия/Италия/Испания), пробел (Финляндия/Франция/Польша/страны Балтии) или апостроф (Швейцария).
- Формат бумаги — Letter (США/Канада/Мексика) или A4 (везде else).

Нажмите «Сохранить», чтобы глобально сохранить настройки локали. В разделе «Импорт данных» также имеется переключатель «Локализация» для отдельных файлов, предназначенный для разовых импортов, в которых используются иные соглашения. В том же раскрывающемся списке «Настройки» находится поле «Корневая папка regProj», которое позволяет изменить расположение общего дерева проектов, используемого mgcvUI, earthUI и glmnetUI.

Совет: Формат чисел на осях графика и в метках наклона автоматически адаптируется к вашей языковой локали. В немецкой локали для разделителя тысяч используется точка (200.000), в финской — пробел (200 000), а в швейцарской — апостроф (200'000). Символы валют не отображаются — mgcvUI не зависит от валюты.

4 Режим оценки

Если цель активного проекта указана как «Для оценки», mgcvUI настраивается на оценку одного объекта недвижимости. Все функции, описанные в главе 3, остаются доступными; в данной главе рассматриваются только дополнительные возможности, специфичные для оценки.

4.1 Обработка строк объекта оценки

В режиме оценки первая строка вашего набора данных представляет собой объект оценки, а все остальные строки — аналог сделки. Ваш вводный файл должен быть организован соответствующим образом (см. главу 2). Цена продажи объекта может быть оставлена пустой или установлена на любое значение — mgcvUI автоматически рассматривает её как NA во время подгонки модели.

После подгонки модель генерирует прогнозы для соответствующей строки объекта оценки, что служит основой для рабочего процесса корректировки RCA.

4.2 Дата оценки & Возраст, с которого разрешена продажа

В режимах «Оценка» и «Рынок» в разделе «Настройка переменных» появляется поле «Дата оценки» (по умолчанию установлена сегодняшняя дата). Если вы выберете для столбца значение «contract_date» в раскрывающемся списке «Special», mgcvUI рассчитает столбец «sale_age» — количество целых дней между датой заключения договора по каждой сделке и датой оценки. Этот столбец добавляется в качестве предиктора.

При изменении даты оценки показатель «sale_age» автоматически пересчитывается.

4.3 Специальный столбец Обозначения

В режимах оценки и рынка для каждого предиктора в таблице «Конфигурация переменных» появляется специальный раскрывающийся список. Полную справку по специальным типам и их эффектам см. в главе 6.

4.4 Обзор корректировок RCA

Кнопка «Рассчитать корректировки RCA» (&) (отображается в режимах оценки и рыночного анализа после подгонки аналогов) позволяет рассчитать рыночные корректировки для каждого аналога по отношению к объекту оценки. Полный рабочий процесс RCA описан в главе 11.

Совет: Балл CQA, присваиваемый объекту оценки, определяет, какая доля остатков распределения относится именно к этому объекту оценки. Балл 5,00 означает, что объект оценки находится на медиане по сравнению с аналогами.

5 Режим анализа рыночной зоны

Если целью активного проекта является «Анализ рыночной зоны», mgcvUI предоставляет те же функции, характерные для сферы недвижимости, что и в режиме оценки (специальные столбцы, срок с момента продажи), но ориентирован на анализ группы объектов недвижимости, а не на оценку отдельного объекта оценки.

5.1 Отличия от режима оценки

- Обработка строк с объектами оценки осуществляется гибко — в режиме «Рынок» строка 1 не рассматривается автоматически как объект оценки.
- Отсутствует таблица продаж — этап «Сгенерировать таблицу продаж& Скачать» недоступен.

5.2 Когда следует использовать режим «Рынок»

Режим «Анализ рыночной зоны» подходит в следующих случаях:

- Построение модели GAM для района или рыночной зоны с целью понимания нелинейных факторов, определяющих стоимость
- Анализ того, как такие переменные, как площадь, возраст, площадь участка и расположение, влияют на цены продажи в группе объектов недвижимости
- Подготовка материалов для анализа рыночной конъюнктуры или района
- Работа с набором данных, в котором требуются специальные характеристики столбцов (срок продажи), но не нужен рабочий процесс с таблицей продаж

Совет: Режим «Рынок» также полезен для общей регрессионного анализа на рынке недвижимости, когда требуются специальные характеристики столбцов, но не нужен рабочий процесс оценки отдельного объекта недвижимости.

6 Отбор переменных

В разделе 4 боковой панели — «Настройка переменных» — вы выбираете, какие столбцы будут включены в модель и как с ними будет проводиться работа.

6.1 Целевая переменная

В раскрывающемся списке «Целевая (реакционная) переменная» в верхней части раздела 4 перечислены все числовые столбцы вашего набора данных. Выберите один столбец в качестве реакционной переменной (например, «Цена продажи»). Целевой столбец автоматически исключается из списка предикторов.

6.2 Преобразование отклика

Под селектором цели в раскрывающемся списке «Преобразование ответа» доступны три варианта:

- Нет (сырые данные) — Использовать ответную переменную в исходном виде. Подходит, когда зависимость между предикторами и ответной переменной является приблизительно аддитивной в исходной шкале.
- Логарифм (естественный) — Применить преобразование по естественному логарифму перед подгонкой. При обратном преобразовании это делает корректировки пропорциональными (%), а не абсолютными (\$). Особенно полезно для цен продажи, где корректировка \$ 10 000 означает совершенно разные вещи при разных ценовых уровнях.
- Log10 — Применить логарифмическое преобразование с основанием 10.

Совет: Логарифмическое преобразование делает корректировки по времени пропорциональными (%), а не абсолютными (\$). Для моделей рынка недвижимости, охватывающих широкий диапазон цен, логарифмическое преобразование часто позволяет получить остатки с более предсказуемым поведением.

При выборе логарифмического преобразования значения ≤ 0 в ответе автоматически отфильтровываются. Все выводные данные (прогнозы, вклады, остатки) автоматически преобразуются обратно в исходную шкалу.

6.3 Таблица предикторов

Ниже селектора целей в прокручиваемой таблице перечислены все остальные столбцы. Для столбцов с флажками в целях компактности используются повернутые вертикальные заголовки. Порядок столбцов следующий:

Столбец	Описание
Variable	Название столбца (моноширинный шрифт, усечено с использованием многоточия). Ячейка с рамкой.
Type	Раскрывающийся список типов данных: числовой, целочисленный, символьный, факторный, логический, Дата, POSIXct. Раскрывающийся список с рамкой.
Include	Флажок — включить этот столбец в модель в качестве предиктора. Вертикальный заголовок.
Factor	Флажок — рассматривать как категориальный фактор (создает отдельный коэффициент для каждого уровня). Вертикальный заголовок.
Linear	Флажок — принудительно использовать линейную зависимость (без сглаживания, только βx). Вертикальный заголовок.
Special	Раскрывающийся список (только для оценки/рынка) — см. Справочник по специальным типам столбцов ниже. Раскрывающийся список с рамкой.
NAs	Количество и процент отсутствующих значений, обозначенных цветом: красный ($\geq 50\%$), оранжевый ($\geq 20\%$), серый (> 0). Ячейка с рамкой.

«Фактор» vs. «Сглаживание» vs. «Линейный»: по умолчанию для включенных числовых переменных создается член сглаживания $f(x)$. Установка флажка «Фактор» приводит к созданию категориального члена (по одному коэффициенту на каждый уровень). Установка флажка «Линейный» создаёт простой линейный член βx вместо процесса сглаживания. Переменные, отмеченные как «Фактор» и «Линейный», обрабатываются как «Фактор».

6.4 Определение типов данных & Переопределения

mgcvUI автоматически определяет типы данных при импорте. Распознаются столбцы с числовыми, целочисленными, логическими, факторными и датными значениями. Вы можете переопределить любое определение, изменив значение в раскрывающемся списке «Тип». Изменение типов влияет на то, как столбец обрабатывается в модели.

6.5 Специальные типы столбцов См.

В режимах «Оценка» и «Рынок» в раскрывающемся списке «Специальное» доступны следующие параметры:

Взвешивание:

- weight — Столбец с весами наблюдений (допускается только одна; строки с весом = 0 исключаются из подгонки)

Дата & . Типы времени:

- contract_date — Запускает автоматический расчет показателя «sale_age» с даты оценки
- sale_age — Обозначает заранее рассчитанный столбец «Срок продажи» (количество дней с даты оценки), используемый без изменений
- listing_date — Используется в качестве резервного варианта для расчета показателя «Days on Market»
- dom — Обозначает столбец «Days on Market»

Тип записи:

- sale_type — Определяет столбец типа записи, который позволяет отличить данный объект оценки от аналогов сделок

Типы денежных единиц:

- `concessions` — Определяет условия скидок при продаже (кредиты продавца, поощрения для покупателя и т. д.)

Размер & . Типы мест:

- `latitude` — Значения автоматически округляются до 3 знаков после запятой
- `longitude` — Применяется тот же метод округления, что и для широты
- `area` — Идентификатор рыночной зоны или района
- `living_area` — Включает расчет остаточной стоимости на квадратный фут (`residual_sf` и `cqa_sf`)
- `lot_size` — Столбец «Площадь участка»
- `site_dimensions` — Группируется с площадью участка

Возрастные группы:

- `actual_age` — Столбец «Возраст объекта недвижимости»
- `effective_age` — Фактический возраст объекта недвижимости

Типы дисплеев:

- `display_only` — Столбец включается в экспортные файлы Excel, но полностью исключается из процесса подгонки модели. Используйте его для полей адреса, номеров MLS, идентификаторов земельных участков или других справочных данных.

Совет: Столбцы, помеченные как «`display_only`» (не используемые в модели), остаются в наборе данных и отображаются при экспорте в Excel, но полностью исключаются из таблицы предикторов. Используйте эту функцию для столбцов с идентификаторами, адресами или другими справочными полями.

7 Выбор параметров

Раздел 5 боковой панели — «Параметры вызова MgcV» — предоставляет доступ ко всем параметрам настройки GAM. У каждого параметра есть синий значок справки (?) с всплывающей подсказкой.

7.1 Предустановки параметров

В раскрывающемся списке в верхней части экрана доступны две предустановки, которые задают оптимальные настройки по умолчанию для типичных рабочих процессов:

Предустановки	Базис	Настройки
Автономный режим (поиск)	tp (тонкая пластина)	k = auto (10), gamma = 1.2, select = TRUE
Earth Pipeline (уточнение)	cr (кубическая регрессия)	k = auto (10), gamma = 1,4, select = TRUE

Пресет «Earth Pipeline» выбирается автоматически при импорте узлов earthUI. Для интеграции узлов earthUI требуется базис кубического регрессионного сплайна (cr), поскольку он позволяет указывать точные координаты узлов.

7.2 Семья

Выберите семейство распределений для вашей зависимой переменной:

Family	Use Case
gaussian	Непрерывные значения (например, цена продажи). Наиболее распространенный случай.
Gamma	Положительные непрерывные значения с дисперсией, пропорциональной среднему значению.
poisson	Данные, выраженные в количестве (например, количество продаж).
binomial	Бинарные исходы.
inverse.gaussian	Данные с сильным правосторонним асимметричным распределением.

7.3 Метод

Метод оценки параметра сглаживания:

Метод	Описание
REML	Ограниченное максимальное правдоподобие (по умолчанию, рекомендуется). Наиболее устойчиво к переобучению.
GCV.Cp	Обобщённая перекрёстная проверка. Имеет тенденцию выбирать несколько более сложные модели.

Метод	Описание
ML	Метод максимального правдоподобия. Аналогичен REML, но может приводить к недостаточному сглаживанию.
P-REML	Вариант REML по Пирсону.
P-ML	Вариант ML по Пирсону.
GACV.Cp	Обобщённая приближённая перекрёстная проверка.

7.4 Gamma (коэффициент сглаживания)

Параметр «gamma» умножает количество эффективных степеней свободы в критерии сглаживания, способствуя получению более гладких (менее «волнистых») подгонок. Значение по умолчанию — 1,2.

Значение	Эффект
1,0	Стандартное сглаживание
1,2–1,4	Немного более плавное, чем по умолчанию (рекомендуется)
2,0+	Значительно более плавное — подходит для данных с шумом или небольших выборок

Более высокие значения параметра gamma предотвращают переобучение за счет более сурового штрафа за сложность модели. В пресете «Earth Pipeline» используется значение 1,4 для дополнительного сглаживания при доработке узлов модели Земли.

7.5 Перекрёстная проверка

Если этот флажок установлен (по умолчанию), mgcvUI после подгонки модели выполняет 10-кратную перекрёстную проверку для вычисления коэффициента R-квадрат по результатам перекрёстной проверки. Это позволяет получить объективную оценку прогностической способности модели на вневывборочной выборке.

7.6 Выбрать (Сжатие)

Если этот флажок установлен (по умолчанию), к каждому члену, подвергающемуся сглаживанию, добавляется дополнительный штраф, который может привести к его сведению к нулю. Это позволяет осуществлять автоматический отбор переменных — незначимые члены, подвергающиеся сглаживанию, фактически удаляются из модели. Реализовано с помощью функции ‘select = TRUE’ в файле ‘gam()’.

7.7 Базовые данные по умолчанию

Тип базисных функций сплайна для сглаживаемых слагаемых:

Базис	Полное имя	Описание
tp	Сплайн регрессии тонкой пластины	По умолчанию. Изотропный (размещение узлов не требуется). Хороший универсальный выбор.
cr	Кубический регрессионный сплайн	Позволяет явно задавать расположение узлов. Требуется для интегрирования по узелкам на поверхности Земли.

Базис	Полное имя	Описание
ps	P-сплайн	Равномерно распределённая В-сплайн-основа с штрафом за разность. Вычислительно эффективна.
bs	B-сплайн	Гибкая В-сплайн-основа.

7.8 Значение k по умолчанию (базисная размерность)

Параметр «basis dimension» k определяет максимальную «волнистость» каждого сглаживания. Значение 0 (по умолчанию) означает «автоматический режим» — mgcvUI использует $k = 10$ или количество «земных узлов», в зависимости от того, что более уместно.

- Небольшое значение k (3–5) : Очень плавные кривые, практически линейные.
- Среднее значение k (8–15) : Умеренная гибкость (как правило, достаточная).
- Большое значение k (20+) : Очень высокая гибкость — существует риск переобучения без сильной регуляризации.

Совет: Показатель «Эффективные степени свободы» (EDF) на вкладке «Сводка» (Summary) показывает, сколько степеней свободы фактически использует каждая функция сглаживания. Если значение EDF близко к $k - 1$, базис, возможно, слишком мал — увеличьте k . Если значение EDF значительно меньше k , функция сглаживания подвергается сильному штрафу.

7.9 Тип тензора

Для взаимодействий между непрерывными переменными доступны два типа тензорного произведения: а именно:

Тип	Функция	Описание
ti	ti(x1, x2)	Тензорное взаимодействие — моделирует только эффект взаимодействия, при этом основные эффекты обрабатываются отдельно с помощью одномерных $s()$ слагаемых. Предпочтительно с точки зрения интерпретируемости.
te	te(x1, x2)	Тензорное произведение — моделирует весь совокупный эффект, включая основные эффекты. Более сложно разложить для корректировок RCA.

7.10 Допустимые взаимодействия

В сворачиваемом разделе «Допустимые взаимодействия» отображается верхнетреугольная матрица флажков для всех включенных сглаживаний (нелинейных, нефакторных) предикторов. Каждый флажок включает тензорное произведение сглаживания между соответствующей парой переменных.

- «Разрешить всё» / «Снять всё» — Кнопки массового переключения.
- Щелкните по имени переменной — Включить или отключить все взаимодействия для этой переменной.

- Щелкните правой кнопкой мыши по имени переменной — «Исключить из основного эффекта» — помечает переменную для использования только в взаимодействиях (без одномерного сглаживания).

При импорте узлов earthUI взаимодействия, обнаруженные earth, предварительно проверяются и блокируются (выделяются желтым фоном). Если в используемой модели earth установлен параметр `degree = 1` (без взаимодействий), отображается информационное сообщение.

Warning

Включение тензорных взаимодействий существенно повышает сложность модели. Используйте взаимодействия только в том случае, если у вас есть предметные знания или данные наземных наблюдений, подтверждающие их обоснованность. Убедитесь, что полученные корректировки являются обоснованными и объяснимыми.

7.11 Взаимодействия «фактор сглаживание»

При включении в модель как факторных, так и переменных сглаживания появляется отдельная матрица взаимодействий «Фактор × Переменная сглаживания». Каждый установленный флажок создает член $s(x, by = factor)$ — отдельную кривую сглаживания для каждого уровня фактора.

7.12 Расширенные параметры

При раскрытии раздела «Дополнительно» появляются дополнительные настройки:

- Оптимизатор — Алгоритм сглаживания оценки параметров: `outer/newton` (по умолчанию), `outer/bfgs` или `efs` (расширенный метод Феллнера-Шалла).
- Масштаб — Параметр масштаба. 0 (по умолчанию) = оценка по данным. Установите известное значение для моделей с фиксированным масштабом.
- Метод дискретных ковариат — Быстрая дискретизированная подгонка для больших наборов данных (`discrete = TRUE`).
- Потоки — Количество параллельных потоков (1–32). Действует только при значении `discrete = TRUE`.

8 Подгонка модели

8.1 Кнопка «Fit»

В разделе 6 боковой панели находится кнопка «Fit Mgcsv GAM Model ». При нажатии на нее модель запускается с текущими настройками.

8.2 Что происходит во время подгонки

При нажатии кнопки «Подгонка» в mgcsvUI:

- 1 . Проверяет конфигурацию — проверяет наличие как минимум одного предиктора и не менее 10 полных строк. Выдает предупреждение, если менее 50% строк являются полными (с указанием столбцов, в которых содержится наибольшее количество значений NA).
- 2 . Подготавливает данные — формирует подмножества, содержащие только столбцы, значимые для модели, удаляет строки с значениями «NA», применяет веса (если они указаны), исключает строку объекта оценки (в режиме оценки), применяет преобразования отклика.
- 3 . Создает формулу GAM — формирует сглаживающие члены $s(x, bs="cg k=10)$ для числовых предикторов, факторные члены для категориальных переменных, линейные члены для переменных, помеченных как «Линейные», и тензорные произведения для выбранных взаимодействий. Узлы Earth встраиваются в качестве явных позиций узлов в базисных членах типа «cg».
- 4 . Обработывает крайние случаи — удаляет константные переменные, преобразует бинарные предикторы в факторы, ограничивает количество « k » до числа уникальных значений минус 1, дополняет «earth knots» квантилями данных, когда требуется больше узлов.
- 5 . Подгонка GAM — вызывает функцию «mgcsv::gam()» с построенной формулой, семейством, методом, гамма-параметром, параметром «select» и другими параметрами.
- 6 . Запускает перекрёстную проверку (если она включена) — 10-кратная перекрёстная проверка с вычислением R-квадрата вне выборки.
- 7 . Обновляет все вкладки с результатами — «Сводка», «Уравнение», графики вклада, «Диагностика» и все остальные вкладки заполняются данными.

8.3 Общая информация Подгонка& Ход работы

Если доступен пакет callr, подгонка модели выполняется в фоновом процессе, благодаря чему приложение продолжает реагировать на запросы пользователя. На экране отображается накладка с индикатором хода выполнения в стиле терминала с темной темой:

- Таймер, отображающий прошедшее время в секундах/минутах.
- Журнал трассировки с цветовой кодировкой: синий — для статуса, желтый — для хода сглаживания блока CV, красный — для ошибок, зеленый — для результатов.
- Кнопка «Прервать» для остановки процесса подгонки в случае необходимости.
- Кнопка «Заккрыть», появляющаяся после завершения.

Если функция callr недоступна, подгонка выполняется синхронно (приложение зависает до завершения подгонки).

После успешной подгонки на кнопке «Fit» появляется белая галочка, а в строке состояния отображается следующая информация: «R-sq = X, CV R-sq = X, Dev = X%, AIC = X, n = X».

9 Вкладки «Результаты»

9.1 Данные

Отображает импортированные данные в виде интерактивной таблицы DataTable с горизонтальной прокруткой и 15 строками на страницу. В режимах «Оценка» и «Рынок» вкладка разделена на таблицу «Объект оценки» (строка 1) и таблицу «Аналоги сделок» (остальные строки); в режиме «Общий» отображается единая таблица.

Каждая ячейка отображается в одной строке и обрезается до ширины столбца, поэтому широкие поля свободного ввода (например, примечания к объектам недвижимости) не растягивают строку вниз по странице. Один щелчок выделяет строку; двойной щелчок по любой ячейке открывает всплывающее окно с её полным, необрезанным содержимым.

9.2 Уравнение

Отображает подогнанную модель GAM в двух частях:

Образец уравнения (отображение с помощью MathJax):

- В левой части показана зависимость с соответствующим преобразованием ($\log$, $\log_{10}$) и функцией связи.
- В правой части показаны значение пересечения, члены сглаживаемые при сглаживании в виде $f_i(\text{variable})$, факторные члены, сгруппированные по переменным с указанием числа уровней, а также линейные члены со значениями коэффициентов.
- Семейство и метод отображаются над уравнением.

Таблица определений функций сглаживания:

- Строка формулы члена (например, $s(\text{living_sqft}, \text{bs} = "cr k=8")$)
- Имя переменной
- Тип (s , te , ti или линейный)
- Тип базисной функции
- Значение k
- Координаты узелков на поверхности Земли (если импортированы)

9.3 Резюме

В верхней части экрана отображаются шесть метрических карточек:

- R-квадрат — доля объясненного отклонения
- CV R-квадрат — R-квадрат перекрёстной проверки (если включена перекрёстная проверка)
- Dev. Explained% — процент объяснённого отклонения
- AIC — информационный критерий Акайке
- n — количество использованных наблюдений
- Smooths — количество сглаживающих слагаемых

Под карточками:

- Таблица сглаженных членов — Член, EDF (эффективные степени свободы), Ref.df, F-статистика, p-значение
- Таблица параметрических членов — Член, оценка, стандартная ошибка, t-значение, p-значение

9.4 Важность переменных

Горизонтальная гистограмма, отражающая относительную значимость каждого члена модели:

- Сглаживание членов : ранжированы по F-статистике (синие столбцы)
- Параметрические члены : ранжированы по значению $|t|$ (зеленые столбцы)

Под диаграммой в таблице DataTable приведены столбцы «Срок», «Тип», «EDF», «Статистика» и «p-значение», отсортированные по значениям статистического показателя в порядке убывания.

9.5 Графики вкладов

Каждый член модели, влияющий на прогнозируемое значение, имеет интерактивный график в Plotly, отображаемый в карточке с рамкой. На вкладке представлены графики для всех типов членов:

9.5.1 Univariate Smooth Terms — $s(x)$

- Синяя линия с лентой доверия 95%
- Серые точки, отображающие частичные остатки (для моделей без логарифмического преобразования)
- Красные пунктирные вертикальные линии в точках расположения «земных узлов» (если «земные узлы» импортированы)
- Всплывающая подсказка при наведении курсора : значение переменной, вклад (\$), 95% CI, коэффициент (наклон на единицу)

В случае моделей с логарифмическим преобразованием ось Y преобразуется обратно в вклады в долларах с помощью формулы $\bar{y} \times (e^{f(x)} - 1)$, что позволяет непосредственно интерпретировать кривые в долларовом выражении.

Для переменных широты и долготы наклон указывается с шагом 0,001 градуса, а не с шагом в один градус, поскольку такой шаг является более значимым географическим шагом.

9.5.2 Фактор-сглаживание Термины — $s(x, \text{by} = \text{factor})$

По одной цветной линии на каждый уровень фактора, наложенные на один график. Полезно для того, чтобы увидеть, как влияние предиктора варьируется между группами (например, между разными районами).

9.5.3 Interaction Terms — $ti(x_1, x_2)$ and $te(x_1, x_2)$

Тепловая карта, отображающая двумерную поверхность вкладов. Цветовая шкала: красный (отрицательный) $\rightarrow$ белый (нулевой) $\rightarrow$ синий (положительный). При наведении курсора отображаются как значения переменных, так и величина вклада.

9.5.4 Параметрические члены — линейные и факторные

- Числовые линейные слагаемые : диаграмма рассеяния точек данных с наложенной линией подгонки
- Факторные слагаемые : гистограмма, показывающая средний вклад по каждому уровню фактора

9.6 Корреляция

Матрица тепловой карты, отображающая корреляции Пирсона между всеми числовыми предикторами. Доступна сразу после импорта данных — подгонка модели не требуется.

- Цветовой градиент: синий (−1) → белый (0) → красный (+1)
- Значения корреляции, выведенные в каждую ячейку
- Полезно для выявления конкурвивности (сглаживания мультиколлинеарности) перед подгонкой

9.7 Диагностика

Два набора диагностических визуализаций:

Диагностическая панель (с сайта `gratia::appraise`): четыре графика — остатки по сравнению с прогнозируемыми значениями, Q-Q-график, гистограмма остатков и зависимость реакции от прогнозируемых значений.

Диаграмма рассеяния «фактические значения против прогнозируемых»: наблюдаемые значения против прогнозируемых с пунктирной линией отсчета под углом 45 градусов. Для моделей с логарифмическим преобразованием значения преобразуются обратно в исходную шкалу.

9.8 Корректировки RCA

Три гистограммы, отображаемые после расчёта RCA (только в режиме «оценка/рынок»):

- Корректировка остаточной стоимости% — Распределение корректировок остаточной стоимости в процентах от цены продажи
- Чистая корректировка% — Распределение чистых корректировок
- Валовая корректировка% — Распределение валовых корректировок

В подписи к каждой гистограмме указаны среднее значение, медиана и стандартное отклонение, а также нанесены пунктирные линии, обозначающие среднее значение и медиану.

9.9 Anova

Таблица комбинированного ANOVA, объединяющая параметрические и члены сглаживания из сводки GAM. В ней указаны член, тип (параметрический или сглаживание) и все соответствующие статистические показатели.

9.10 Вывод Mgcsv:

Вывод необработанного текста:

- Подгонка времени
- Формула GAM
- Вывод результатов моделирования
- Полный вывод результатов по методу «`summary(model)`»
- Семейство и функция связи
- Общие показатели конкурвивности

9.11 Проверка знаков

При импорте узлов `earthUI` на этой вкладке сравнивается направление каждой переменной в модели Earth с её направлением в GAM:

- `earth _` направление : возрастающее, убывающее или смешанное (на основе знаков функции шарнира)

- `gam _` направление : возрастающее, убывающее или смешанное (оценивается по гладкой производной на таблице из 100 точек)
- согласованность : TRUE/FALSE
- предупреждение : описание любого несоответствия

В строке состояния отображается сообщение «Все направления согласуются» или количество переменных с несоответствиями.

9.12 Concurvity

Конкурвита — это «сглаживающийся» аналог мультиколлинеарности. На этой вкладке отображаются:

- Таблица общей кривизны : строки для показателей «наихудший», «наблюдаемый» и «оценочный», столбцы — для каждого члена сглаживания. Цветовая кодировка: белый ($< 0,5$), жёлтый ($0,5-0,8$), красный ($> 0,8$).
- Таблица парной конкурвити : матрица парных значений для наихудшего случая с такой же цветовой кодировкой.

Значения, превышающие $\sim 0,8$ в строке «worst», указывают на то, что данное сглаживание может быть хорошо аппроксимировано другими сглаживаниями, что свидетельствует о избыточности.

9.13 Проверка исходных данных

Вывод, полученный с помощью команды `mgcv::gam.check()`: проверка адекватности размерности базиса для каждого члена сглаживания. Если показатель адекватности базиса (k) члена сглаживания слишком мал, эта проверка помечает его низким p -значением, что указывает на необходимость увеличить коэффициент сглаживания (k).

10 Загрузка данных

После подгонки кнопка «Скачать результаты (Excel)» (боковая панель, раздел 7) позволяет экспортировать файл Excel с прогнозами и данными для диагностики в папку вывода mgcv активного проекта.

10.1 Столбцы вывода

Столбец	Описание
basis	Значение переменной «пересечение» после обратного преобразования (одинаковое для всех строк)
<var>_contribution	Вклад каждого сглаживаемого члена в прогноз. Для логарифмических моделей пропорционально распределяется по значениям в долларах.
est_<target>	Прогноз модели (например, est_sale_price), обратно преобразованный из логарифмического масштаба, если применимо
residual	Фактическое значение— прогнозируемое значение
calc_residual	Верификация: фактическое значение— сумма вкладов (обратно преобразованная)
cqa	Оценка «Состояние-Качество-Привлекательность» (CQA) (шкала от 0 до 10)
residual_sf	Остаточная / жилая площадь (если обозначено как « living_area »)
cqa_sf	Показатель CQA, рассчитанный на основе рейтинга по адресу residual_sf

В режиме оценки в строке 1 (объект оценки) значения sf для остаточных величин/cqa/cqa _ установлены как NA. Строки сортируются по значениям sf остаточных величин _ в порядке убывания, если указан столбец жилой площади _.

После успешного завершения на кнопке загрузки появляется белая галочка.

10.2 Результаты CQA: балл

балл

CQA оценивает остаток каждой строки по сравнению со всеми остальными по шкале от 0 до 10:

- Высокий показатель CQA ($\sim 9-10$) : продан значительно дороже прогнозируемой цены — вероятно, отличное состояние/качество/привлекательность
- Низкий показатель CQA ($\sim 0-1$) : продан значительно дешевле прогнозируемой цены — вероятно, неудовлетворительное состояние или продажа в связи с финансовыми затруднениями
- Показатель CQA ~ 5 : близок к медиане

11 Расчеты RCA& Скачивание

Рабочий процесс Метода остаточных ограничений доступен в режимах «Оценка» и «Анализ рыночной зоны» после подгонки модели. Объект оценки должен находиться в первой строке.

11.1 Открытие диалогового окна RCA

Нажмите кнопку «Рассчитать корректировки RCA» & «Скачать» в разделе 8 боковой панели. Появится модальное диалоговое окно со следующим содержанием:

- Тип балла : CQA или CQA на SF (если указано « living_area »). Вариант «на SF» масштабирует остаточный балл объекта оценки с учетом жилой площади.
- Балл CQA объекта : Принимает значения от 0,00 до 10,00. Это поле изначально пустое; после импорта модели earthUI тип бала и значение, использованные в earthUI, переносятся из общих настроек проекта.

Нажмите «Generate» для вычисления и загрузки. Файл Excel сохраняется в папке вывода mgcv активного проекта (<os>_out_mgcgv).

11.2 Интерполяция баллов CQA

- 1 . Балы CQA и остатки для аналогов отсортированы
- 2 . Линейная интерполяция (approx()) сопоставляет значение CQA с остатком
- 3 . Значение для объекта оценки = прогноз модели + интерполированный остаток
- 4 . Строки с весом 0 обрабатываются так, как если бы у них был тот же остаток, что и у объекта оценки

11.3 Столбцы вывода

Колонка	Описание
subject_value	Прогноз по модели + интерполированное отклонение
subject_cqa	Оценка CQA, введенная пользователем
<var>_adjustment	Вклад субъекта– Вклад компании (в долларах)
residual_adjustment	Остаток субъекта– Остаток компании
net_adjustments	Сумма всех корректировок
gross_adjustments	Сумма абсолютных корректировок
residual_adj_pct	Корректировка остатка как% от цены продажи
net_adj_pct	Чистые корректировки как% от цены продажи
gross_adj_pct	Валовые корректировки как% от цены продажи
adjusted_sale_price	Фактическая цена продажи + чистые корректировки

После успешного выполнения вычислений на кнопке появляется белая галочка.

12 Таблица сравнения продаж

Таблица сравнения продаж доступна только в режиме оценки, после расчёта корректировок RCA. Она формирует отформатированную книгу Excel, подходящую для включения в отчёты об оценке.

12.1 Диалоговое окно выбора аналогов

При нажатии кнопки «Сгенерировать таблицу продаж» & «Скачать» открывается модальное диалоговое окно со списком всех аналогов. Аналоги разделены на две группы:

- Рекомендуемые аналоги (уже отмечены): Валовая корректировка < 25% от цены продажи, отсортированные по дате продажи (сначала самые последние). Максимум 30 аналогов.
- Другие аналоги (по умолчанию не отмечены): Остальные аналоги, отсортированные по проценту валовой корректировки.

Выберите аналоги, а затем нажмите «Создать таблицу продаж», чтобы создать книгу.

Обязательные специальные типы : `contract_date` и `living_area` должны быть указаны. Рекомендуемые специальные типы : `latitude` , `longitude` и `lot_size` — таблица будет работать и без них, но некоторые поля останутся пустыми.

12.2 Макет книги

Сгенерированная книга содержит несколько листов, на каждом из которых размещены по 3 аналога рядом друг с другом:

- Строки заголовков : адреса объекта оценки и покупателя, цены продажи, даты продажи и количество дней на рынке
- Сгруппированные строки : местоположение (широта, долгота, площадь), размер участка (площадь участка, габариты участка) и возраст (фактический возраст, эффективный возраст) с заголовками групп и подпунктами
- Строки с переменными модели : по одной строке на каждый сглаживающий член, показывающие вклад объекта оценки, вклад аналогов и суммы корректировок
- Строки с остаточными характеристиками : пустые, разблокированные ячейки для ввода оценщиком для ввода характеристик, не учтенных моделью (например, состояние, качество, вид)
- Сводные строки : чистая корректировка% , валовая корректировка% и скорректированная цена продажи с активными формулами Excel

12.3 Защита листа & Остатки ячейки

Каждый лист защищен от случайного редактирования, однако ячейки со значениями остаточных характеристик явно разблокированы. Это позволяет оценщикам вводить значения для характеристик, отсутствующих в модели, при этом сохраняя расчеты корректировок, основанные на формулах.

13 Скачивание отчетов

Создание отчетов представляет собой двухэтапный рабочий процесс в Quarto, заменивший одноэтапный процесс создания отчетов, использовавшийся в предыдущих версиях.

13.1 Шаг 1: Создание отчёта Quarto Отчёт «

»

Кнопка «Generate Quarto Отчёт» (доступная после того, как модель прошла подгонку) сохраняет автономный набор «.qmd» в папку вывода mgsu активного проекта. Набор содержит исходный код Quarto, предварительно сгенерированные файлы графиков (PNG и PDF), файл `report_data.rds` с данными, включающий все таблицы и метаданные модели, а также шаблон стиля `reference.docx` для вывода в Word. Поскольку набор является автономным, вы можете редактировать исходный код .qmd или объединять его с другими документами Quarto перед выводом.

Графики предварительно рендерятся при генерации набора; если генерация отдельного графика завершается сбоем, записывается изображение-заполнитель, а остальная часть набора генерируется в обычном режиме.

13.2 Шаг 2: Конвертация Quarto Отчёт «

»

Раздел «Convert Quarto Report» преобразует файл «.qmd» в один или несколько форматов вывода с помощью «`quarto::quarto_render()`». По умолчанию используется последний созданный набор, но кнопка «Browse» позволяет выбрать любой файл «.qmd». Доступны три формата:

- HTML — Автономный HTML-документ. Работает на всех платформах без дополнительного программного обеспечения. Рекомендуемый вариант по умолчанию.
- Word (.docx) — Оформляется с использованием входящего в набор шаблона стиля «reference.docx». Подходит для редактирования и распространения.
- PDF — Требуется установки LaTeX (см. «Системные требования» ниже). Если LaTeX не обнаружен, опция PDF автоматически скрывается из списка флажков форматов.

Устаревшие одношаговые рендереры — `render_gam_report()` и `export_gam_docx()` — по-прежнему доступны для программного использования.

13.3 Содержание отчёта

Сгенерированный отчёт содержит:

- Набор данных & Модель : файл данных, размер выборки, семейство, метод, а также информация о подгонке
- Сводка результатов : R^2 , $CV R^2$, объясненное отклонение, AIC и связанные показатели подгонки
- Уравнение модели : уравнение модели с использованием нотации f_i для сглаживающих слагаемых и явных коэффициентов для линейных слагаемых, а также определения сглаживающих функций. Длинные уравнения разбиваются на несколько строк в PDF.
- Сводные таблицы : процесс сглаживания (EDF, F, p-значение) и параметрические члены (оценка, стандартная ошибка, t, p-значение)
- Включенные взаимодействия : Матрица тензорных и факторно-сглаженных взаимодействий, включенных в модель (отображается только при наличии взаимодействий)
- Важность переменных : Гистограмма, ранжированная по F-статистике (процесс сглаживания) или по коэффициенту « $|t|$ » (параметрические члены)

- Сглаженные эффекты (все типы членов):
 - Одномерные графики с сглаживанием кривых (статические версии интерактивных графиков, созданные с помощью ggplot2)
 - Многолинейные графики «фактор × сглаживание кривой»
 - Тепловые карты взаимодействий (термины *ti* и *te*)
 - Графики параметрических слагаемых (точечный график для числовых, гистограмма для факторов)
- Тепловая карта корреляций предикторов
- Диагностика : диагностика остатков и рассеяние фактических значений по отношению к прогнозируемым
- mgcv Сводка : Полный вывод результатов подгонки модели
- Приложение с обозначениями : Непронумерованный заключительный раздел, объясняющий обозначения функций вклада, использованные в отчёте

После успешного создания отчёта на кнопке «Создать отчёт Quarto» появляется белая галочка.

14 Экспорт функций модели

mgcvUI позволяет экспортировать функции, подвергшиеся сглаживанию при подгонке GAM-модели, в виде автономного кода на нескольких языках программирования. Каждая функция, подвергшаяся сглаживанию, вычисляется на таблице из 200 точек и экспортируется в виде таблицы поиска с линейной интерполяцией, что позволяет использовать модель вне среды R без какой-либо зависимости от пакета mgcv.

В текущей версии экспорт функций осуществляется программным способом. Экспортируемые вспомогательные функции принимают объект результатов подгонки (созданный с помощью `fit_gam()`) и записывают файлы, описанные ниже:

- `generate_r_code()`, `generate_python_code()` и `generate_cpp_code()` — исходные файлы для каждого языка
- `export_functions_sqlite()` — база данных SQLite с таблицами поиска
- `export_functions_zip()` — формирует набор из любой комбинации четырёх форматов в zip-архив

14.1 Функции R

Файл: `gam_functions.R`

- Комментарии заголовка с переменной отклика, R^2 , $CV R^2$ и формулой.
- Константа `INTERCEPT`.
- По одному `approxfun()` на каждую сглаживаемую переменную с заранее вычисленными векторами x/y .
- Коэффициенты линейных слагаемых в виде именованных констант.
- Использование: сложить пересечение, каждый `g_var(value)` и каждый `coef_var * value`.

14.2 Функции Python

Файл: `gam_functions.py`

- `numpy import`.
- Константа `INTERCEPT`.
- Для каждой переменной: массивы `_x_var` и `_fx_var`, функция `g_var(x)` с использованием `np.interp()`.
- Коэффициенты линейных слагаемых в виде констант `COEF_*`.

14.3 Заголовочный файл C++

Файл: `gam_functions.hpp`

- Только заголовок с использованием `#pragma once`.
- Пространство имён `gam_functions`.
- Вспомогательная функция `interp()` с использованием `std::lower_bound`.
- Для каждой переменной: встроенная функция со статическими массивами констант.
- Коэффициенты линейных слагаемых в виде `constexpr double`.

14.4 База данных SQLite

Файл: `gam_functions.sqlite`

- Таблица `smooth_grids`: столбцы (`variable`, `x`, `fx`) — таблицы поиска по 200 точкам.
- Таблица `model_info`: отклик, r^2 в квадрате, `cv_rsqr`, отклонение, объясненная доля, `aic`, `n` — количество наблюдений, семейство, метод, формула.

- Таблица `linear_terms` : переменная, коэффициент.
- Таблица `intercept` : значение пересечения.

15 Сравнение с earthUI и glmnetUI

Приложение «Earth» и его модули постобработки представляют собой единый набор инструментов для регрессионного моделирования. Они используют общий формат данных, специальные типы столбцов, рабочий процесс RCA и демонстрационные наборы данных, но при этом используют разные механизмы моделирования.

15.1 Основные различия

Функции	mgcvUI	earthUI	glmnetUI
Методы	GAM (mgcv)	MARS/Earth (earth)	эластичная сеть (glmnet)
Взаимосвязи	Сглаживаемые нелинейные	Кусочно-линейные	Линейные
Отбор переменных	select=TRUE Сжатие	Пошаговое прореживание	Штраф Lasso
Взаимодействия	Тензорные произведения (te/ti)	Шарнирные произведения (степень 2–3)	Парные произведения
Частичные эффекты	Гладкие кривые с доверительным интервалом	g-функции (кусочные)	Линейные наклоны
Конвейер Earth	Импорт узлов из Earth	Источник узлов	Импорт базиса шарниров
Экспорт кода	R, Python, C++, SQLite	R, Python, C++, SQLite	—
Форматы отчетов	Word, PDF, HTML	Word, PDF, HTML	Word, PDF, HTML

Стоит отметить одну отличительную особенность, присущую исключительно earthUI: earthUI включает экспериментальную функцию Prolog/DCG, которая извлекает структурированные характеристики объектов недвижимости из текстовых комментариев MLS.

15.2 Общие функции

Все три инструмента предоставляют:

- Общее дерево проектов regProj (по умолчанию ~/regProj) — проекты, файлы ввода и выходы для каждого метода используются совместно всеми приложениями одного уровня
- Импорт из CSV и Excel с помощью snake_ Преобразование столбцов с учетом регистра
- Поддержка локалей более 30 стран с указанием разделителя в CSV, знака десятичной запятой и формата бумаги
- Специальные обозначения столбцов (контракт_ дата, жилая площадь_ , площадь, широта, долгота и т. д.)
- Автоматический расчет срока действия продажи_ расчета срока действия с даты оценки
- Переключение между темным и светлым режимами (тема Nord)
- Сохранение настроек (SQLite и/или localStorage)
- Тепловая карта корреляции, важность переменных, диагностика
- Оценка CQA и рабочий процесс корректировки RCA

- Таблица сравнения продаж с выбором аналогов для сравнения и отформатированным выводом в Excel
- Совместимость с теми же демонстрационными наборами данных (Appraisal_1.csv)

15.3 Рекомендуемый рабочий процесс

- Начните с earthUI , чтобы выявить важные переменные, нелинейные зависимости и взаимодействия. Автоматическое построение модели в earthUI — лучшая отправная точка.
- Если вам нужны более плавные частичные эффекты сглаживания, доработайте модель с помощью mgcvUI . Импортируйте узлы из earthUI для плавного перехода.
- Используйте glmnetUI, если вам нужна чисто линейная модель с агрессивным отбором переменных или ограничениями на знаки коэффициентов.
- Сравните все три подхода на одних и тех же данных, чтобы оценить, оправдана ли дополнительная сложность нелинейных моделей.

16 Демо-набор данных: Оценка _ 1.csv

16.1 Описание

Демо-набор данных используется как в earthUI, так и в glmnetUI. Он находится в папке «demo_mls/» исходного кода mgcvUI, либо, если установлен earthUI, в:

```
demo_file <- system.file("extdata", "Appraisal_1.csv", package = "earthUI")
```

Файл содержит примерно 1 500 записей о продажах жилой недвижимости, полученных в результате симулированного экспорта данных из системы MLS, причем рассматриваемый объект оценки находится в первой строке, а аналогичные сделки — в остальных строках. Данные отражают продажи односемейных домов на рынке, охватывающем несколько районов, с различными параметрами площади, возраста и местоположения объектов недвижимости.

Это не реальные данные, но они основаны на реальном районе в Северной Калифорнии. Вся идентифицирующая информация была изменена или удалена.

16.2 Столбцы

Колонка	Тип	Специальный тип	Описание
sale_type	символ	sale_type	Код типа записи, позволяющий отличить данный объект от аналог продаж
weight	числовой	weight	Вес наблюдения (0 = исключить из моделирования)
id	числовой	display_only	Внутренний идентификатор записи
property_id	числовой	display_only	Идентификатор объекта недвижимости в системе MLS
listing_id	символ	display_only	Номер объявления в системе MLS
parcel_number	символ	display_only	Номер участка по оценке округа (APN)
street_address	символ	display_only	Адрес объекта недвижимости
city_name	символ	display_only	Город
postal_code	символ	display_only	Почтовый индекс
county_name	символ	display_only	Округ
contract_date	Дата	contract_date	Дата заключения договора купли- (рассчитывается sale_age)
sale_age	числовой	—	Количество дней от даты заключения договора до даты оценки (рассчитано заранее)
coe_date	Дата	display_only	Дата завершения эскроу
listing_status	символьный	display_only	Статус объявления (например, «Продано»)
sale_price	числовой	(целевой)	Цена продажи — переменная ответа
rent	числовой	—	Ежемесячная арендная плата (для моделей с несколькими целевыми переменными)
list_price	числовой	display_only	Цена объявления
original_list_price	числовой	display_only	Первоначальная цена объявления
living_sqft	числовой	living_area	Общая жилая площадь в квадратных футах
ppsf	числовой	display_only	Цена за квадратный фут
beds_total	целое число	—	Количество спален
baths_total	числовое	—	Общее количество ванных комнат (например, 2,5 = 2 полные + 1 туалет)
lot_size	числовое	lot_size	Площадь участка в квадратных футах
area_id	целое число	area	Идентификатор района в системе MLS
area_text	символ	display_only	Название района
age	числовое	actual_age	Возраст объекта недвижимости в годах

Колонка	Тип	Специальный тип	Описание
year_built	целое число	display_only	Год постройки
latitude	числовое	latitude	Широта (округлённая до 3 знаков после запятой для модели)
longitude	числовой	longitude	Долгота (округлённая до 3 знаков после запятой для модели)
latitude6	числовой	display_only	Широта с полной точностью
longitude6	числовой	display_only	Долгота с полной точностью
garage_spaces	целое число	—	Количество гаражных мест
fp_count	целое число	—	Количество каминов
no_of_stories	число	—	Количество этажей
style	символ	—	Архитектурный стиль
view	символ	—	Тип вида (например, «Окрестности», «Холмы»)
days_on_market	целое число	dom	Количество дней на рынке
listing_date	Дата	listing_date	Дата размещения объявления
sale_concessions	число	concessions	Уступки продавца

16.3 Рекомендуемое руководство по быстрому запуску

- 1 . Запустите процедуру `mgcv: earthUI::launch_mgcv()`
- 2 . Создайте новый проект (раздел 1) с целью «For оценка»
- 3 . Поместите файл `Appraisal_1.csv` в папку вводных данных проекта и выберите его в разделе 2 (при необходимости нажмите «Refresh»)
- 4 . В качестве целевого показателя выберите « sale_price »
- 5 . При желании установите для параметра «Response Transform» значение «Log (natural)» для пропорциональных корректировок
- 6 . Назначьте специальные типы, как показано в таблице выше
- 7 . Включить предикторы: `sale_age` , `living_sqft` , `baths_total` , `lot_size` , `area_id` (в качестве фактора), `age` , `latitude` , `longitude` , `garage_spaces`
- 8 . Сохранить настройки по умолчанию: `REML`, `gamma = 1.2`, `tp basis`, `k = auto`, `select = TRUE`
- 9 . Нажмите «Подгонка `Mgcv GAM Model`»
- 10 . Просмотрите вкладки «Summary», вкладки «Contribution» (графики сглаживания) и «Диагностика»
- 11 . Скачайте вывод (шаг 7), просмотрите рейтинг `CQA`
- 12 . Рассчитайте корректировки `RCA` (шаг 8) с баллом `CQA ~5,00`
- 13 . Сгенерируйте таблицу сравнения продаж (шаг 9) с рекомендуемыми аналогами

17 Системные требования& Устранение неполадок

17.1 Поддерживаемые платформы

mgcvUI работает на macOS, Windows и Linux. Приложение разработано и в основном протестировано на macOS с использованием RStudio. Примечания, касающиеся конкретных платформ, приведены ниже.

17.2 Версия R:

Требуется $R \geq$ версии 4.1.0. Настоятельно рекомендуется использовать RStudio Desktop (версия 2023.06 или более поздняя) — в него входит набор пакетов Quarto и pandoc (необходимые для визуализации отчетов), а также он предоставляет удобную среду для запуска приложения.

17.3 Необходимые пакеты R:

Следующие пакеты устанавливаются автоматически в качестве зависимостей:

Package	Purpose
shiny	Фреймворк для веб-приложений
mgcv	Подгонка моделей GAM
valengrCore	Общий модуль оценки (специальные роли, перенос настроек из earthUI)
ggplot2, plotly	Статические и интерактивные графики вклада
gratia	Извлечение и диагностика сглаживаемых членов
DT	Интерактивные таблицы данных
quarto, rmarkdown	Создание и визуализация отчётов с помощью Quarto
readxl	Импорт файлов Excel
shinyFiles	Браузеры файлов и папок
RSQLite, DBI	Настройки и базы данных проектов (SQLite)
jsonlite	Сериализация состояния переменных

17.4 Дополнительные зависимости

Component	When Needed
LaTeX (TinyTeX, MiKTeX или MacTeX)	Только для генерации отчетов в формате PDF. Если установка LaTeX не обнаружена, опция PDF автоматически скрывается из раскрывающегося списка форматов. Установить с помощью: <code>tinytex::install_tinytex()</code>
sysfonts + showtext	Шрифт Roboto Condensed для графиков ggplot2. Если он недоступен или отсутствует, приложение автоматически переключается на системный шрифт sans-serif.
earth ($\geq 5.3.0$)	Требуется только для конвейера импорта earthUI. Не требуется для автономного использования.
callr	Фоновая подгонка модели (обеспечивает отзывчивость приложения). Без него подгонка выполняется синхронно.
writexl, openxlsx	Загрузка файлов Excel (вывод моделирования, RCA, таблица продаж). Приложение выдает запрос, если они отсутствуют.
bslib	Тематическое оформление Bootstrap 5 (Nord Light/Dark).
officer	Устаревший программный экспорт в Word (<code>export_gam_docx()</code>).

17.5 Примечания по конкретным платформам

17.5.1 macOS

Минимальное количество проблем. Для системных библиотек рекомендуется использовать Homebrew. Если требуются отчёты в формате PDF:

```
# In R:
install.packages("tinytex")
tinytex::install_tinytex()
```

17.5.2 Windows

Хорошо работает с RStudio Desktop. Важные замечания:

- Корпоративные/заблокированные компьютеры : если доступ к временному каталогу или каталогу пользовательских данных ограничен, сохранение настроек и формирование отчётов могут завершиться сбоем. Приложение отобразит предупреждение в консоли R, но продолжит работу без сохранения настроек.
- LaTeX : Установите TinyTeX из R (та же команда, что и для macOS выше) или установите MiKTeX с сайта miktex.org .
- Длинные пути к файлам : В Windows действует ограничение на длину пути в 260 символов. Старайтесь, чтобы путь к рабочему каталогу был коротким.

17.5.3 Linux

Наибольшие различия наблюдаются между различными дистрибутивами. Пользователям Ubuntu/Debian могут потребоваться системные библиотеки для компиляции пакетов R:

```
sudo apt install libcurl4-openssl-dev libssl-dev \
libxml2-dev libsqlite3-dev libfontconfig1-dev
```

Для просмотра отчетов в формате PDF: `tinytex::install_tinytex()` или установите `texlive-full` через системный менеджер пакетов.

На серверах без монитора (без дисплея) интерактивные графики Plotly отображаются в отчётах в виде статических запасных вариантов. Для работы самого приложения Shiny требуется подключение к веб-браузеру.

17.6 Плавная деградация

`mgcvUI` разработан таким образом, что может работать даже при отсутствии дополнительных компонентов:

Отсутствующий компонент	Поведение
Отсутствует LaTeX	Опция PDF скрыта в раскрывающемся списке форматов отчетов. Форматы HTML и Word по-прежнему доступны.
Отсутствует интернет / ошибка шрифтов	Шрифт Roboto Condensed заменен системным шрифтом sans-serif. В консоль занесено сообщение.
Отсутствует RSQLite / файловая система только для чтения	Настройки не сохраняются между сеансами. Отображается предупреждение в консоли. В остальном приложение работает нормально.
<code>gratia::appraise</code> завершается с ошибкой	На вкладке «Диагностика» и в отчетах отображается сообщение-заполнитель. В качестве альтернативы используйте команду <code>mgcv::gam.check()</code> в консоли.
Временный каталог недоступен для записи	Сгенерировать отчёт не удастся с явным сообщением об ошибке. Устраните проблему, установив переменную среды <code>TMPDIR</code> в путь к каталогу, доступному для записи.

17.7 Устранение неполадок в программе «

»

17.7.1 «Параметр PDF недоступен»

Установка LaTeX не обнаружена. Выполнить в R:

```
install.packages("tinytex")
tinytex::install_tinytex()
```

Затем перезапустите приложение. Появится опция «PDF».

17.7.2 «Настройки не сохраняются»

Не удаётся создать каталог базы данных SQLite или выполнить запись в него. Проверьте права доступа к:

- macOS/Linux : `~/Library/Application Support/R/mgcvUI/` или `~/local/share/R/mgcvUI/`
- Windows : `%APPDATA%/R/data/mgcvUI/`

17.7.3 «График диагностики недоступен»

Функция `gratia::appraise()` иногда завершается с ошибкой при работе со сложными моделями (множество тензорных взаимодействий, нестандартные семейства). Это известная проблема в исходном коде. Обходной путь: запустите команду `mgcv::gam.check(model)` в консоли R, чтобы сразу увидеть графики диагностики.

17.7.4 “Отчёт об ошибке: количество элементов, которые необходимо заменить ...”

Это была известная проблема в более ранних версиях, связанная с пакетом `gratia`. Убедитесь, что у вас установлена последняя версия `mgcvUI`. Текущая версия перехватывает эту ошибку и генерирует отчёт с заполнителем в разделе диагностики.

17.7.5 Порт 7880 уже занят

Если приложение не запускается из-за того, что порт 7880 занят (в результате предыдущего сеанса), выполните следующую команду:

```
# macOS/Linux:
lsof -ti:7880 | xargs kill
```

```
# Windows (PowerShell):
Stop-Process -Id (Get-NetTCPConnection -LocalPort 7880).OwningProcess
```

17.7.6 Ошибка установки пакета в Linux

Отсутствуют системные библиотеки. Установите заголовочные файлы для разработки, перечисленные в разделе «Linux» выше, а затем повторите попытку выполнить команду `install.packages()`.

Список литературы

- Wood, S. N. (2003). «Thin-plate regression splines». B: Journal of the Royal Statistical Society (B) 65.1, с. 95—114. DOI: [10.1111/1467-9868.00374](https://doi.org/10.1111/1467-9868.00374).
- (2004). «Stable and efficient multiple smoothing parameter estimation for generalized additive models». B: Journal of the American Statistical Association 99.467, с. 673—686. DOI: [10.1198/016214504000000980](https://doi.org/10.1198/016214504000000980).

- Wood, S. N. (2011). «Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models». В: Journal of the Royal Statistical Society (B) 73.1, с. 3—36. DOI: [10.1111/j.1467-9868.2010.00749.x](https://doi.org/10.1111/j.1467-9868.2010.00749.x).
- (2017). Generalized Additive Models: An Introduction with R. 2-е изд. Chapman и Hall/CRC.
- Wood, S. N., N. Pya и B. Säfken (2016). «Smoothing parameter and model selection for general smooth models (with discussion)». В: Journal of the American Statistical Association 111, с. 1548—1575. DOI: [10.1080/01621459.2016.1180986](https://doi.org/10.1080/01621459.2016.1180986).