

# mgcvUI: Eine interaktive Benutzeroberfläche für das mgcv-Paket (generalisierte additive Modelle)

*Deutsche Übersetzung, mit KI-Unterstützung erstellt und redaktionell geprüft; die englische*

*Fassung ist maßgeblich.*

William Bert Craytor

Jahrgang 1, Ausgabe 1 · Sommer 2026

Veröffentlicht am 1. Juli 2026

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

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

Dieser Open-Access-Artikel wird unter den Bedingungen der [Creative-Commons-Namensnennung-4.0-Lizenz](#) verbreitet.

**Zusammenfassung.** mgcvUI ist eine grafische Benutzeroberfläche für das R-Paket mgcv, das generalisierte additive Modelle (GAMs) mit penalisierten Regressionssplines und automatischer Glättungsparameterwahl anpasst. Es bietet drei Verwendungsmodi — allgemeine prädiktive Modellierung, Immobilienbewertung und Marktgebietsanalyse — und führt den Anwender durch Datenimport, Spezifikation der Glättungsterme, Modellanpassung, Diagnose- und Effektdiagramme sowie herunterladbare Berichte. Dieser Artikel dokumentiert die Datenformat-Anforderungen, den Modellierungs-Workflow, die Ausgabedarstellungen und die vollständige Funktionsreferenz von mgcvUI.

**Schlüsselwörter:** generalisierte additive Modelle, GAM, penalisierte Splines, Glättung, mgcv, R, grafische Benutzeroberfläche, Immobilienbewertung

**Anmerkung des Verlags (Juli 2026):** Kurz vor Drucklegung wurde mgcvUI in earthUI (Version 0.11.0 und höher) zusammengeführt; earthUI stellt den hier beschriebenen GAM-Workflow nun neben den MARS- und Elastic-Net-Workflows bereit. mgcvUI bleibt in seiner letzten veröffentlichten Fassung (0.3.0) unter <https://github.com/wcraytor/mgcvUI> verfügbar; die Weiterentwicklung erfolgt in earthUI. Dieser Artikel bezieht sich auf mgcvUI in der veröffentlichten Form.

## Inhaltsverzeichnis

|          |                                                     |          |
|----------|-----------------------------------------------------|----------|
| <b>1</b> | <b>Einleitung</b>                                   | <b>4</b> |
| 1.1      | Geschichte und Hintergrund                          | 4        |
| 1.1.1    | Generalisierte additive Modelle                     | 4        |
| 1.1.2    | Das Paket mgcv                                      | 4        |
| 1.1.3    | Vergleich mit MARS (earth) und Elastic Net (glmnet) | 5        |
| 1.1.4    | Die earthUI–mgcvUI-Pipeline                         | 6        |
| 1.1.5    | Empfohlener Arbeitsablauf                           | 6        |
| 1.2      | Was ist mgcvUI?                                     | 6        |
| 1.3      | Die drei Purpose-Modi                               | 7        |
| 1.4      | Immobilien-spezifische Funktionen                   | 7        |
| 1.5      | Erste Schritte                                      | 8        |
| <b>2</b> | <b>Anforderungen an MLS-Eingabedaten</b>            | <b>9</b> |
| 2.1      | Dateiformat & Struktur                              | 9        |
| 2.2      | Erforderliche Spalten im Wertermittlungsmodus       | 9        |
| 2.3      | Konventionen für besondere Spaltenbezeichnungen     | 10       |

|          |                                                                        |           |
|----------|------------------------------------------------------------------------|-----------|
| 2.4      | Datenqualität & Vollständigkeit                                        | 10        |
| 2.5      | Platzierung der Bewertungsobjektzeile                                  | 10        |
| <b>3</b> | <b>Der Modus General Purpose</b>                                       | <b>11</b> |
| 3.1      | Überblick                                                              | 11        |
| 3.2      | Der Arbeitsablauf in der Seitenleiste                                  | 11        |
| 3.3      | Import aus earthUI (optional)                                          | 12        |
| 3.4      | Registerkarten des Hauptbereichs                                       | 12        |
| 3.5      | Dauerhafte Speicherung der Einstellungen                               | 13        |
| 3.6      | Dark Mode                                                              | 13        |
| 3.7      | Locale & Regional Settings (Gebietsschema und regionale Einstellungen) | 13        |
| <b>4</b> | <b>Der Appraisal-Modus</b>                                             | <b>15</b> |
| 4.1      | Behandlung der Bewertungsobjektzeile                                   | 15        |
| 4.2      | Effective Date & Sale Age                                              | 15        |
| 4.3      | Besondere Spaltenkennzeichnungen                                       | 15        |
| 4.4      | Überblick über die RCA-Anpassungen                                     | 15        |
| <b>5</b> | <b>Der Modus Market Area Analysis</b>                                  | <b>16</b> |
| 5.1      | Unterschiede zum Appraisal-Modus                                       | 16        |
| 5.2      | Wann der Market-Modus sinnvoll ist                                     | 16        |
| <b>6</b> | <b>Variablenauswahl</b>                                                | <b>17</b> |
| 6.1      | Zielvariable                                                           | 17        |
| 6.2      | Transformation der Zielgröße                                           | 17        |
| 6.3      | Die Prädiktortabelle                                                   | 17        |
| 6.4      | Erkennung und Überschreiben von Datentypen                             | 18        |
| 6.5      | Referenz der Special-Spaltentypen                                      | 18        |
| <b>7</b> | <b>Parameterwahl</b>                                                   | <b>20</b> |
| 7.1      | Parameter-Voreinstellungen                                             | 20        |
| 7.2      | Family (Verteilungsfamilie)                                            | 20        |
| 7.3      | Method (Schätzverfahren)                                               | 20        |
| 7.4      | Gamma (Glättungsmultiplikator)                                         | 21        |
| 7.5      | Kreuzvalidierung                                                       | 21        |
| 7.6      | Select (Schrumpfung)                                                   | 21        |
| 7.7      | Default Basis (Standardbasis)                                          | 21        |
| 7.8      | Default k (Basisdimension)                                             | 21        |
| 7.9      | Tensortyp                                                              | 22        |
| 7.10     | Zugelassene Interaktionen                                              | 22        |
| 7.11     | Faktor-Glättungsterm-Interaktionen                                     | 22        |
| 7.12     | Erweiterte Parameter                                                   | 22        |
| <b>8</b> | <b>Anpassen des Modells</b>                                            | <b>24</b> |
| 8.1      | Die Schaltfläche Fit                                                   | 24        |
| 8.2      | Was während der Anpassung geschieht                                    | 24        |
| 8.3      | Anpassung im Hintergrund & Fortschritt                                 | 24        |
| <b>9</b> | <b>Ergebnis-Registerkarten</b>                                         | <b>25</b> |
| 9.1      | Data                                                                   | 25        |
| 9.2      | Equation                                                               | 25        |
| 9.3      | Summary                                                                | 25        |
| 9.4      | Variable Importance                                                    | 25        |

|           |                                                           |           |
|-----------|-----------------------------------------------------------|-----------|
| 9.5       | Beitragsdiagramme                                         | 26        |
| 9.5.1     | Univariate Glättungsterme — $s(x)$                        | 26        |
| 9.5.2     | Faktor-Glättungsterme — $s(x, \text{by} = \text{Faktor})$ | 26        |
| 9.5.3     | Interaktionsterme — $ti(x_1, x_2)$ und $te(x_1, x_2)$     | 26        |
| 9.5.4     | Parametrische Terme — linear und faktoriell               | 26        |
| 9.6       | Correlation                                               | 26        |
| 9.7       | Diagnostics                                               | 27        |
| 9.8       | RCA Adjustments                                           | 27        |
| 9.9       | Anova                                                     | 27        |
| 9.10      | Mgcv Output                                               | 27        |
| 9.11      | Sign Check                                                | 27        |
| 9.12      | Concurvity                                                | 28        |
| 9.13      | Basis Check                                               | 28        |
| <b>10</b> | <b>Herunterladen der Daten</b>                            | <b>29</b> |
| 10.1      | Ausgabespalten                                            | 29        |
| 10.2      | CQA-Werte                                                 | 29        |
| <b>11</b> | <b>RCA-Berechnungen und Download</b>                      | <b>30</b> |
| 11.1      | Öffnen des RCA-Dialogs                                    | 30        |
| 11.2      | Interpolation des CQA-Werts                               | 30        |
| 11.3      | Ausgabespalten                                            | 30        |
| <b>12</b> | <b>Vergleichswertraster (Sales Comparison Grid)</b>       | <b>31</b> |
| 12.1      | Dialog zur Auswahl der Vergleichsobjekte                  | 31        |
| 12.2      | Aufbau der Arbeitsmappe                                   | 31        |
| 12.3      | Blattschutz & Zellen für Restmerkmale                     | 31        |
| <b>13</b> | <b>Herunterladen von Berichten</b>                        | <b>32</b> |
| 13.1      | Schritt 1: Generate Quarto Report                         | 32        |
| 13.2      | Schritt 2: Convert Quarto Report                          | 32        |
| 13.3      | Inhalt des Berichts                                       | 32        |
| <b>14</b> | <b>Export von Modellfunktionen</b>                        | <b>34</b> |
| 14.1      | R-Funktionen                                              | 34        |
| 14.2      | Python-Funktionen                                         | 34        |
| 14.3      | C++-Header                                                | 34        |
| 14.4      | SQLite-Datenbank                                          | 34        |
| <b>15</b> | <b>Vergleich mit earthUI und glmnetUI</b>                 | <b>36</b> |
| 15.1      | Wesentliche Unterschiede                                  | 36        |
| 15.2      | Gemeinsame Funktionen                                     | 36        |
| 15.3      | Empfohlener Arbeitsablauf                                 | 36        |
| <b>16</b> | <b>Demodatensatz: Appraisal_1.csv</b>                     | <b>38</b> |
| 16.1      | Beschreibung                                              | 38        |
| 16.2      | Spalten                                                   | 38        |
| 16.3      | Vorgeschlagener Schnelleinstieg                           | 39        |

|                                                      |           |
|------------------------------------------------------|-----------|
| <b>17 Systemanforderungen &amp; Fehlerbehebung</b>   | <b>40</b> |
| 17.1 Unterstützte Plattformen                        | 40        |
| 17.2 R-Version                                       | 40        |
| 17.3 Erforderliche R-Pakete                          | 40        |
| 17.4 Optionale Abhängigkeiten                        | 41        |
| 17.5 Plattformspezifische Hinweise                   | 41        |
| 17.5.1 macOS                                         | 41        |
| 17.5.2 Windows                                       | 41        |
| 17.5.3 Linux                                         | 41        |
| 17.6 Verhalten bei fehlenden Komponenten             | 42        |
| 17.7 Fehlerbehebung                                  | 42        |
| 17.7.1 „PDF option is not available“                 | 42        |
| 17.7.2 „Settings will not persist“                   | 42        |
| 17.7.3 „Diagnostics plot unavailable“                | 43        |
| 17.7.4 „Report error: number of items to replace...“ | 43        |
| 17.7.5 Port 7880 bereits belegt                      | 43        |
| 17.7.6 Paketinstallation schlägt unter Linux fehl    | 43        |

## 1 Einleitung

### 1.1 Geschichte und Hintergrund

#### 1.1.1 Generalisierte additive Modelle

**Generalisierte additive Modelle** (Generalized Additive Models, GAMs) wurden 1990 von Trevor Hastie und Robert Tibshirani in ihrer Monographie formalisiert. GAMs erweitern das lineare Modell, indem sie jeden linearen Term  $\beta_j x_j$  durch eine glatte Funktion  $f_j(x_j)$  ersetzen; das Modell lautet dann:

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

Jedes  $f_j$  wird mithilfe eines penalisierten Regressionssplines aus den Daten geschätzt. Die glatten Funktionen erlauben es dem Modell, nichtlineare Zusammenhänge abzubilden, ohne dass der Anwender vorab eine funktionale Form festlegen muss. Der Strafterm steuert die Welligkeit jedes Glättungsterms und verhindert eine Überanpassung, lässt aber genügend Flexibilität, um tatsächliche Muster nachzuzeichnen.

#### 1.1.2 Das Paket mgcv

Das R-Paket `mgcv` (Mixed GAM Computation Vehicle) wurde von Simon Wood an der University of Bath entwickelt (Wood 2017). Erstmals 2001 veröffentlicht, ist es heute die R-Standardimplementierung von GAMs und Bestandteil jeder R-Installation. Die zugrunde liegenden Verfahren sind in einer Reihe von Arbeiten beschrieben, die die stabile Schätzung mehrerer Glättungsparameter (Wood 2004), Thin-Plate-Regressionssplines (Wood 2003), die schnelle stabile REML-Schätzung (Wood 2011) sowie die allgemeine Auswahl glatter Modelle (Wood, Pya und Säfken 2016) behandeln. Zu den wesentlichen Merkmalen zählen:

- **Automatische Glättungsparameterwahl:** Der Glättungsparameter jedes Terms wird im Rahmen der Modellanpassung geschätzt — mittels REML (Restricted Maximum Likelihood), GCV (Generalized Cross-Validation, generalisierte Kreuzvalidierung) oder ML (Maximum Likelihood).

- **Verschiedene Basistypen:** Thin-Plate-Regressionssplines (tp), kubische Regressionssplines (cr), P-Splines (ps) und B-Splines (bs) — jeweils mit unterschiedlichen Eigenschaften.
- **Tensorprodukt-Glättungsterme:** `te()` und `ti()` ermöglichen glatte Interaktionen zwischen Variablen, ohne dem Fluch der Dimensionalität zu unterliegen.
- **Variablenselektion:** Die Option `select = TRUE` fügt einen zusätzlichen Strafterm hinzu, der Glättungsterme auf null schrumpfen kann und damit faktisch eine Variablenselektion vornimmt.
- **Concurvity-Diagnostik:** Misst das Analogon zur Multikollinearität für Glättungsterme.

### 1.1.3 Vergleich mit MARS (earth) und Elastic Net (glmnet)

Die Anwendung `earth` und ihre beiden Post-Processing-Module (`glmnet` und `mgcv`) verwenden unterschiedliche Modellierungs-Engines. Die folgende Tabelle fasst die wesentlichen Unterschiede zusammen:

| Merkmal                            | mgcv (GAM)                                                                 | earth (MARS)                                                                 | glmnet (Elastic Net)                                           |
|------------------------------------|----------------------------------------------------------------------------|------------------------------------------------------------------------------|----------------------------------------------------------------|
| <b>Modelltyp</b>                   | Glatt nichtlinear (penalisierte Splines)                                   | Stückweise linear (adaptive Splines)                                         | Linear (mit Regularisierung)                                   |
| <b>Nichtlinearität</b>             | Automatisch über glatte Funktionen (Thin Plate, kubisch usw.)              | Automatisch über Hinge-Funktionen mit datengetriebenen Knoten                | Nur über Interaktionen oder Basiserweiterung                   |
| <b>Variablenselektion</b>          | Optional über doppelten Schrumpfung-Strafterm ( <code>select=TRUE</code> ) | Integriert über schrittweise Vorwärts-/Rückwärtsauswahl mit GCV-Beschneidung | Integriert über $L_1$ -Strafterm (Lasso)                       |
| <b>Interaktionen</b>               | Tensorprodukt-Glättungsterme ( <code>te()</code> , <code>ti()</code> )     | Hinge-basierte Produkte (interpretierbar, bis Grad 3)                        | Paarweise Kreuzprodukte (schwer interpretierbar)               |
| <b>Interpretierbarkeit</b>         | Glatte partielle Effektkurven; sehr anschaulich                            | g-Funktionen zeigen stückweise lineare partielle Effekte je Variable         | Koeffizienten sind lineare Gewichte; einzeln leicht erklärbar  |
| <b>Am besten geeignet für</b>      | Glatte nichtlineare Zusammenhänge, flexible Modellierung                   | Automatische Erkennung von Nichtlinearität und Interaktionen                 | Hochdimensionale Daten, Variablenselektion, kleine Stichproben |
| <b>RCA-Anpassungen</b>             | Glatte Anpassungen über partielle Effekte                                  | Stückweise lineare Anpassungen über g-Funktionen                             | Lineare Anpassungen je Variable                                |
| <b>Kontrolle der Überanpassung</b> | Schätzung des Glättungsparameters über REML/GCV                            | GCV-basierte Beschneidung                                                    | Kreuzvalidiertes Lambda                                        |

**MARS** (Multivariate Adaptive Regression Splines) wurde 1991 von Jerome Friedman eingeführt. Das Verfahren bildet stückweise lineare Modelle, indem es Hinge-Funktionen und deren Knotenpositionen adaptiv aus den Daten auswählt. Die R-Implementierung ist das Paket `earth` von Stephen Milborrow.

Die **Elastic-Net**-Regression, implementiert im Paket **glmnet** von Friedman, Hastie und Tibshirani (2010), verbindet Lasso- ( $L_1$ ) und Ridge-Strafterme ( $L_2$ ) zur gleichzeitigen Variablen-selektion und Koeffizientenschrumpfung.

#### 1.1.4 Die earthUI–mgcvUI-Pipeline

Ein leistungsfähiger Workflow verbindet die automatische Knotenfindung von earthUI mit der Glättungsschätzung von mgcvUI. Wenn earthUI eine Ergebnisdatei im Format `.rds` exportiert und mgcvUI diese importiert, gilt:

1. **Earth findet die Knoten** — MARS ermittelt datengetriebene Hinge-Positionen (Wendepunkte) für jede Variable.
2. **mgcvUI nutzt sie als Ankerpunkte** — Die earth-Knoten bilden die Basis für kubische Regressionssplines (Basis `cr`) im GAM und verschaffen den Glättungstermen einen Vorsprung durch Knotenpositionen, die tatsächliche Strukturen in den Daten widerspiegeln.
3. **mgcv verfeinert die Anpassung** — Die penalisierte Spline-Schätzung glättet die stückweise lineare earth-Anpassung und erzeugt glatte partielle Effektkurven, während die von earth entdeckte Grundform erhalten bleibt.

Diese Pipeline verbindet explorative Modellierung (earth) mit konfirmatorischer Modellierung (GAM) und vereint die Stärken beider Ansätze.

#### 1.1.5 Empfohlener Arbeitsablauf

Für die Immobilienbewertung und vergleichbare Anwendungen, die interpretierbare und begründbare Modelle erfordern:

1. **Beginnen Sie mit earthUI**, um die wichtigen Variablen, nichtlinearen Zusammenhänge und Interaktionen in Ihren Daten aufzuspüren. Die automatische Basisauswahl und die g-Funktionen von earth liefern ausgezeichnete erste Einsichten.
2. **Verfeinern Sie mit mgcvUI**, wenn Sie glattere partielle Effekte wünschen. Importieren Sie die Knoten von earthUI in mgcvUI, um nahtlos von stückweise linearen zu glatten Modellen überzugehen.
3. **Verwenden Sie glmnetUI**, wenn Sie eine konsequente Variablenselektion (Lasso) benötigen, wenn die Zahl der Prädiktoren die Zahl der Beobachtungen erreicht oder übersteigt, oder wenn Sie aus Gründen der Begründbarkeit ein rein lineares Modell mit Regularisierung wünschen.

### 1.2 Was ist mgcvUI?

mgcvUI ist eine grafische Benutzeroberfläche für das R-Paket **mgcv**. Sie läuft als lokale Shiny-Anwendung — ohne Anmeldung, ohne Server und ohne Benutzerkonten. Sie starten sie aus R heraus, importieren einen Datensatz (CSV oder Excel), konfigurieren Ihr Modell und passen interaktiv ein GAM an.

Seit der aktuellen Version ist mgcvUI kein separat installiertes Paket mehr: Es ist *im* Paket earthUI als dessen GAM-Post-Processing-Modul enthalten und wird mit `earthUI::launch_mgcv()` gestartet. Die Anwendung selbst — und alles, was in diesem Artikel beschrieben wird — ist unverändert; lediglich Installation und Startbefehl unterscheiden sich. Die drei Routinen (earth, glmnet, mgcv) teilen sich einen Projektbaum, eine Einstellungsdatenbank und eine gemeinsame unterstützende Infrastruktur.

Die Anwendung bietet einen vollständigen Arbeitsablauf: Datenimport, Variablenkonfiguration mit Spezifikation der Glättungsterme, Modellanpassung mit Hintergrundverarbeitung, Diagnosedigramme, glatte partielle Effektkurven und herunterladbare Berichte im Word-, PDF- oder HTML-Format.

**Generalisierte additive Modelle** ersetzen den linearen Zusammenhang  $\beta x$  für jeden Prädiktor durch eine glatte Funktion  $f(x)$ . Das bedeutet:

- **Keine Notwendigkeit, funktionale Formen vorzugeben** — die Daten bestimmen die Gestalt jedes Zusammenhangs.
- **Glatte partielle Effekte** — der Beitrag jedes Prädiktors ist eine glatte Kurve statt einer Geraden, sodass leicht erkennbar wird, wie eine Variable die Zielgröße bei verschiedenen Werten beeinflusst.
- **Automatische Glättungsparameterwahl** — `mgcv` bestimmt über REML oder GCV, wie wellig jede Kurve sein soll, und wägt dabei Anpassungsgüte gegen Überanpassung ab.

### 1.3 Die drei Purpose-Modi

Jedes Projekt besitzt einen **Purpose** (Verwendungszweck), der bei der Anlage in Abschnitt 1 der Seitenleiste (**Project**) gewählt und danach festgeschrieben wird. Der Purpose des aktiven Projekts bestimmt, welche Werkzeuge und Oberflächenelemente erscheinen; um in einem anderen Modus zu arbeiten, legen Sie ein Projekt mit dem entsprechenden Purpose an oder öffnen ein solches. Die drei Modi sind:

- **General** (allgemein) — GAM-Regression für beliebige Grundgesamtheiten oder Datensätze. Dies ist der Standardmodus. Er stellt den vollständigen GAM-Workflow ohne fachspezifische Ergänzungen bereit.
- **For Appraisal** (für die Wertermittlung) — GAM-Regression, zugeschnitten auf die Immobilienbewertung. Ergänzt Funktionen für die Einzelobjektbewertung, darunter die Behandlung des Bewertungsobjekts, besondere Spaltenkennzeichnungen und den Residual Constraint Approach (RCA).
- **Market Area Analysis** (Marktgebietsanalyse) — GAM-Regression, zugeschnitten auf Marktgebietsstudien. Ergänzt Funktionen zur Analyse von Objektgruppen in einem definierten Markt, mit optionaler Behandlung einer Bewertungsobjektzeile.

In allen drei Modi ist die eigentliche Modellierungs-Engine identisch — Sie passen stets ein GAM über `mgcv::gam()` an. Die Purpose-Einstellung steuert lediglich, welche zusätzlichen Werkzeuge und Oberflächenelemente verfügbar sind.

### 1.4 Immobilienspezifische Funktionen

Ist entweder **For Appraisal** oder **Market Area Analysis** gewählt, aktiviert `mgcvUI` mehrere Funktionen, die für die Immobilienanalyse konzipiert sind:

- **Besondere Spaltenkennzeichnungen** — Jeder Prädiktor kann mit einer besonderen Rolle versehen werden, etwa `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions` oder `display_only`. Diese Kennzeichnungen steuern, wie die Spalte bei der Anpassung und in der Ausgabe behandelt wird.
- **Spalte Sale Age** — Ist eine Spalte als `contract_date` gekennzeichnet und ein Effective Date (Wertermittlungstichtag) angegeben, berechnet `mgcvUI` eine Spalte `sale_age` (Tage zwischen Verkaufsdatum und Stichtag) und setzt sie als Prädiktor ein.
- **RCA-Berechnungen (nur Wertermittlung)** — Im Wertermittlungsmodus kann `mgcvUI` nach der Modellanpassung die Ausgabe des Residual Constraint Approach (RCA) berechnen. Diese liefert Anpassungen je Vergleichsobjekt, Zusammenfassungen der Netto- und Bruttoanpassungen sowie einen angepassten Verkaufspreis für das Bewertungsobjekt.

**Hinweis:** Der Purpose-Modus General eignet sich für beliebige Datensätze — aus Finanzwesen, Wissenschaft, Ingenieurwesen oder Immobilienwirtschaft. Die Modi für Wertermittlung und Marktanalyse fügen lediglich Komfortfunktionen für Immobilienfachleute hinzu.

## 1.5 Erste Schritte

So verwenden Sie mgcvUI:

1. **Installieren Sie earthUI** in R: `install.packages("earthUI")` — die mgcv-Routine ist enthalten. (Zur Entwicklungsversion und zu optionalen Komponenten siehe „Installing earthUI and vProlog“ im earthUI-Artikel dieser Ausgabe.)
2. **Starten Sie die Anwendung:** Führen Sie in R `earthUI::launch_mgcv()` aus. Die Anwendung öffnet sich in Ihrem Webbrowser auf Port 7880.
3. **Legen Sie ein Projekt an oder öffnen Sie eines** in Abschnitt 1 der Seitenleiste. Ein Projekt legt den Purpose (General, For Appraisal oder Market Area Analysis) sowie den Arbeitsort (Land, Bundesland, Kreis, Stadt) und seine Ein- und Ausgabeordner fest. Projekte werden unter einem gemeinsamen `regProj`-Stammordner abgelegt und mit den Schwesteranwendungen earthUI und glmnetUI geteilt. Das Öffnen eines Projekts mit einem anderen Purpose ist der Weg, den Modus zu wechseln.
4. **Importieren Sie Ihre Daten** in Abschnitt 2, indem Sie eine CSV- oder Excel-Datei aus dem Eingabeordner des aktiven Projekts auswählen. Legen Sie Dateien in diesem Ordner ab und klicken Sie dann auf **Refresh**, um sie aufzulisten.
5. **Import aus earthUI (optional)** — Abschnitt 3 erlaubt den Import einer earthUI-Ergebnisdatei (`.rds`), um das GAM mit von earth gefundenen Knoten zu initialisieren. Überspringen Sie diesen Schritt, wenn Sie kein earthUI-Ergebnis haben.
6. **Konfigurieren Sie die Variablen** — wählen Sie Zielgröße und Prädiktoren, legen Sie Datentypen fest, weisen Sie besondere Spaltenrollen zu und markieren Sie Variablen als Faktor oder linear.
7. **Setzen Sie die mgcv-Parameter** — family, method, gamma, Basistyp, k, Tensortyp, Interaktionen und weitere Optionen.
8. **Passen Sie das Modell an** — klicken Sie auf „Fit Mgcv GAM Model“ und prüfen Sie die Ergebnisse im Hauptbereich.
9. **Exportieren Sie** — laden Sie Prognosen als Excel-Datei herunter, erzeugen und konvertieren Sie einen Quarto-Bericht oder berechnen Sie (im Wertermittlungsmodus) RCA-Anpassungen und ein Vergleichswertraster.

Die Einstellungen werden automatisch in einer SQLite-Datenbank gespeichert und beim erneuten Laden derselben Eingabedatei wiederhergestellt.

## 2 Anforderungen an MLS-Eingabedaten

Bei Arbeitsabläufen der Immobilienbewertung und Marktanalyse stammen Ihre Eingabedaten in der Regel aus einem Export eines Multiple Listing Service (MLS). Dieses Kapitel beschreibt die erwartete Dateistruktur und die Spalten, die mgcvUI verwenden kann.

### 2.1 Dateiformat & Struktur

mgcvUI akzeptiert **CSV**- und **Excel**-Dateien (**.xlsx**, **.xls**). Beim Import werden Spaltennamen automatisch in **snake\_case** umgewandelt — aus „Living SqFt“ wird beispielsweise **living\_sqft**, aus „Contract Date“ wird **contract\_date** und aus „Sale Price“ wird **sale\_price**. Diese Normalisierung sorgt für einheitliche Spaltenbezüge im gesamten Arbeitsablauf. Das beim Import verwendete CSV-Trennzeichen und Dezimaltrennzeichen richtet sich nach den Locale-Einstellungen (siehe Kapitel 3, „Locale & Regional Settings“).

Ihre Datendatei sollte eine flache Tabelle mit einer Zeile je Objekt und einer Spalte je Merkmal sein. Die erste Zeile der Datei muss die Spaltenüberschriften enthalten.

### 2.2 Erforderliche Spalten im Wertermittlungsmodus

mgcvUI arbeitet zwar mit beliebigen Spaltensätzen, der vollständige Wertermittlungs-Workflow profitiert jedoch von den folgenden Spalten:

| Spalte           | Special Type           | Zweck                                                                              |
|------------------|------------------------|------------------------------------------------------------------------------------|
| Sale Price       | (Zielgröße)            | Zielvariable des Modells                                                           |
| Contract Date    | <b>contract_date</b>   | Dient zur Berechnung von <b>sale_age</b> (Tage ab Stichtag)                        |
| Listing Date     | <b>listing_date</b>    | Zusammen mit dem Vertragsdatum zur DOM-Berechnung, falls keine DOM-Spalte vorliegt |
| Days on Market   | <b>dom</b>             | Vermarktungsdauer in Tagen; wird in Exporten angezeigt                             |
| Concessions      | <b>concessions</b>     | Verkaufszugeständnisse; Netto-Verkaufspreis = Verkaufspreis – Zugeständnisse       |
| Living Area (SF) | <b>living_area</b>     | Ermöglicht Residuen je Flächeneinheit ( <b>residual_sf</b> , <b>cqa_sf</b> )       |
| Lot Size         | <b>lot_size</b>        | Spalte für die Grundstücksgröße                                                    |
| Site Dimensions  | <b>site_dimensions</b> | Wird mit der Grundstücksgröße gruppiert                                            |
| Latitude         | <b>latitude</b>        | Auf 3 Nachkommastellen gerundet                                                    |
| Longitude        | <b>longitude</b>       | Auf 3 Nachkommastellen gerundet                                                    |
| Area ID          | <b>area</b>            | Kennung des Marktgebiets bzw. der Nachbarschaft                                    |
| Actual Age       | <b>actual_age</b>      | Tatsächliches Alter des Objekts                                                    |
| Effective Age    | <b>effective_age</b>   | Effektives Alter des Objekts                                                       |
| Address          | <b>display_only</b>    | Erscheint in Exporten; vom Modell ausgeschlossen                                   |

Die Spaltennamen in der Tabellenkalkulation dürfen in einer beliebigen Sprache verfasst sein — die „Special“-Bezeichnungen sind englisch, damit das R-Programm sie besonders behandeln kann. Ansonsten erscheinen die von Ihnen vergebenen Spaltennamen in den Regressionsmodellen, den Grafiken und — bei Wertermittlungen — in den Ausgabeberichten.

Nicht alle Spalten sind erforderlich. mgcvUI passt sich an: Fehlt eine Spalte, entfällt schlicht die zugehörige Funktion. Für Immobilienpreismodelle sind bestimmte Spalten jedoch dringend zu empfehlen:

1. **Sale Age** — die Anzahl der Tage zwischen dem Kaufvertragsdatum und dem Stichtag. Wird eine mehrjährige Verkaufshistorie verwendet, spielt **sale\_age** häufig eine zentrale Rolle.
2. **Living Area** (Wohnfläche) — auch unter Bezeichnungen wie „Living Sqft“ oder „GLA“ (gross living area) geführt.
3. **Total Bath Count** — die Gesamtzahl der ganzen, viertel, halben und dreiviertel Badezimmer.
4. **Garage Bays** oder **Garage Area** — Garagenstellplätze bzw. Garagenfläche.
5. **Lot Size** — die Grundstücksfläche des Objekts.
6. **Longitude, Latitude** und, sofern verfügbar, **Area ID** zur Abbildung geografischer Preisunterschiede.

## 2.3 Konventionen für besondere Spaltenbezeichnungen

mgcvUI erkennt Spalten anhand ihrer **Special-Type-Kennzeichnung**, nicht anhand ihres Spaltennamens. Sie können Ihre Spalten im MLS-Export beliebig benennen — entscheidend ist, dass Sie in der Tabelle Variable Configuration (Kapitel 6) den richtigen Special Type zuweisen.

## 2.4 Datenqualität & Vollständigkeit

- **Fehlende Werte (NA)**: Zeilen mit NA-Werten in einer Prädiktor- oder Zielspalte werden vor der Anpassung automatisch entfernt. mgcvUI beschränkt die Daten auf die im Modell verwendeten Spalten, sodass NAs in nicht verwendeten Spalten keinen Zeilenverlust verursachen.
- **Numerische Spalten**: Verkaufspreis, Wohnfläche, Grundstücksgröße und vergleichbare Felder müssen numerisch sein.
- **Faktorspalten**: Text-, TRUE/FALSE- und R-Faktor-Spalten werden automatisch als kategorial behandelt — sie können nur als Faktoren in das Modell eingehen. Eine **numerische** Spalte (etwa ein numerischer Gebiets- oder Stilcode) wird nur dann zum Faktor, wenn Sie das zugehörige Kontrollkästchen **Factor** setzen; mgcvUI leitet niemals aus dem Wertebereich einer numerischen Variable ab, dass sie kategorial sei. Diskrete numerische Prädiktoren wie **bath\_count** bleiben daher stetig (geglättet). Die Special-Rolle ist von der Faktorentscheidung unabhängig.

***Hinweis:** Prüfen Sie nach dem Import die NA-Spalte in der Prädiktortabelle. Spalten mit vielen fehlenden Werten können dazu führen, dass Zeilen entfallen. Spalten mit  $\geq 50\%$  NAs werden rot markiert und automatisch ausgeschlossen.*

## 2.5 Platzierung der Bewertungsobjektzeile

Im Modus **Appraisal** muss **Zeile 1** das **Bewertungsobjekt** enthalten. Alle übrigen Zeilen sind Vergleichsverkäufe. Die Zeile des Bewertungsobjekts wird von der Modellanpassung ausgeschlossen. Nach der Anpassung erzeugt das Modell dennoch Prognosen für diese Zeile.

Im Modus **Market Area Analysis** ist die Platzierung des Bewertungsobjekts in Zeile 1 optional.

Im Modus **General** gibt es keine besondere Zeilenbehandlung — alle Zeilen werden gleich behandelt.

## 3 Der Modus General Purpose

### 3.1 Überblick

Der Modus General Purpose ist beim Start von mgcvUI voreingestellt. Er stellt den vollständigen GAM-Workflow für beliebige Datensätze bereit — nicht nur für Immobilien. Sie können mgcvUI für wissenschaftliche Daten, Finanzanalysen, ingenieurtechnische Untersuchungen oder jedes Regressionsproblem einsetzen, bei dem glatte nichtlineare Zusammenhänge zu erwarten sind.

Im General-Modus lässt die Oberfläche die immobilienpezifischen Funktionen (Special-Spalten, Sale Age, RCA) weg. Die Seitenleiste ist gestrafft und konzentriert sich auf Variablenauswahl, Parameterkonfiguration, Modellanpassung und Export.

### 3.2 Der Arbeitsablauf in der Seitenleiste

Die Seitenleiste ist in nummerierte, aufklappbare Abschnitte gegliedert, die Sie von der Projektauswahl bis zum Berichtsexport führen. Die meisten Abschnitte erscheinen erst, wenn ein aktives Projekt geöffnet ist, und einige sind vom Purpose des Projekts abhängig. Der Ausgabeordner wird aus dem aktiven Projekt abgeleitet — ein separates Feld für den Ausgabeordner gibt es nicht.

**1. Project** — Projekt anlegen oder öffnen. Ein Projekt legt den Purpose und den Arbeitsort (Land, Bundesland, Kreis, Stadt) sowie seinen Eingabeordner (`<os>_in/`) und Ausgabeordner fest. Projekte werden unter einem gemeinsamen `regProj`-Stammordner abgelegt (Vorgabe `~/regProj`, überschreibbar über die Umgebungsvariable `REGPROJ_ROOT` oder das Settings-Menü) und mit den Schwesteranwendungen `earthUI` und `glmnetUI` geteilt, sodass Modelle aus allen dreien nebeneinander liegen. Klicken Sie auf **Close Project**, um zu wechseln.

**2. Import Data** — Wählen Sie eine CSV- oder Excel-Datei aus dem Eingabeordner des aktiven Projekts. Legen Sie Dateien über den Datei-Explorer bzw. Finder im Ordner ab und klicken Sie dann auf **Refresh**, um sie aufzulisten. Bei Excel-Dateien mit mehreren Blättern erscheint eine Blattauswahl. Spaltennamen werden automatisch in `snake_case` umgewandelt.

**3. Import from earthUI (optional)** — Importieren Sie eine `earthUI`-Ergebnisdatei (`.rds`), um das GAM mit von `earth` gefundenen Knotenpositionen (und mit Variablenauswahlen) zu initialisieren. Dieser Schritt ist optional — überspringen Sie ihn, wenn Sie kein `earthUI`-Ergebnis zum Importieren haben.

**4. Variable Configuration** — Auswahl der Zielvariable, Transformation der Zielgröße (none, log, log10), Prädiktortabelle mit Kontrollkästchen für Include, Factor und Linear. Die Spalte Special erscheint nur in den Modi Appraisal und Market. Einzelheiten siehe Kapitel 6.

**5. MgcV Call Parameters** — Die gesamte Modellkonfiguration: Parameter-Voreinstellungen, family, method, gamma, Kreuzvalidierung, select, Basistyp, k, Tensortyp, Interaktionsmatrix und erweiterte Parameter. Die vollständige Parameterreferenz finden Sie in Kapitel 7.

**6. Fit MgcV GAM Model** — Die Schaltfläche, die das Modell ausführt. Die Ergebnisse erscheinen in den Registerkarten rechts.

**7. Download Output** — Exportiert Prognosen, Residuen, CQA-Werte und Beiträge je Variable als Excel-Datei in den Ausgabeordner des Projekts. In allen Purpose-Modi verfügbar, sobald das Modell angepasst ist.

**8. Calculate RCA Adjustments & Download** — Berechnet die RCA-Anpassungen je Variable für jedes Vergleichsobjekt und interpoliert das Residuum des Bewertungsobjekts über CQA. Erscheint nur in den Modi Appraisal und Market, nach der Anpassung.

**9. Generate Sales Grid & Download** — Erzeugt das Vergleichswertraster (Sales Comparison Grid) und empfiehlt Vergleichsobjekte nach dem Prozentsatz der Bruttoanpassung. Erscheint nur im Appraisal-Modus.

**10. Generate Quarto Report** — Schreibt ein in sich geschlossenes Quarto-Berichtspaket (`.qmd`) — mit allen Diagrammen und Tabellen — in den Ausgabeordner des Projekts. Verfügbar, sobald das Modell angepasst ist.

**11. Convert Quarto Report** — Rendert das erzeugte `.qmd`-Paket nach HTML, Word oder PDF. Für PDF ist eine LaTeX-Installation erforderlich; wird keine erkannt, ist die PDF-Option ausgeblendet. Siehe Kapitel 13.

Die purpose-spezifischen Schritte entfallen (und die übrigen Schritte werden neu nummeriert) in den anderen Modi: Im Market-Modus fehlt der Schritt Sales Grid, im General-Modus fehlen sowohl der RCA- als auch der Sales-Grid-Schritt.

### 3.3 Import aus earthUI (optional)

Abschnitt 3 der Seitenleiste — **Import from earthUI** — erlaubt den Import einer earthUI-Ergebnisdatei (`.rds`). Damit wird die earth-mgcv-Pipeline möglich: Die datengetriebenen Knotenpositionen von earth werden zu Ankerpunkten für die GAM-Glättungsterme.

Wird ein earthUI-Ergebnis importiert, geschieht in mgcvUI Folgendes:

- Es erscheint eine Zusammenfassung mit Zielvariable, Bestimmtheitsmaß, Anzahl der Terme und Knotenpositionen je Variable.
- Die Parameter-Voreinstellung „Earth Pipeline“ wird automatisch gewählt (Basis kubischer Regressionsspline,  $\gamma = 1.4$ ).
- Die Zielvariable wird auf die Zielgröße von earth gesetzt.
- Include wird für die Prädiktorvariablen von earth vorab aktiviert.
- Factor wird für die kategorialen Variablen von earth und Linear für die linearen Prädiktoren von earth vorab aktiviert.
- Von earth erkannte Interaktionen werden vorbelegt (in der Interaktionsmatrix gelb hervorgehoben und gegen manuelle Änderungen gesperrt).
- Eine Schaltfläche **Export Knots CSV** stellt die Knoten von earth als CSV-Datei zum Download bereit.

Die Tabelle Variable Configuration zeigt daraufhin einen Hinweis, dass die Variablenauswahlen (Include, Linear, Factor) und die Spline-Knoten aus dem earth-Modell *übernommen* wurden. Anders als in glmnetUI, wo die glmnet-Anpassung an die Basis von earth gebunden ist, passt mgcvUI das GAM anhand Ihrer aktuellen Auswahl an — earth verschafft lediglich einen Vorsprung, und jedes Bedienelement bleibt änderbar.

Eine Schaltfläche **Clear** entfernt die importierten earth-Daten und setzt auf den eigenständigen Betrieb zurück.

### 3.4 Registerkarten des Hauptbereichs

Nach dem Datenimport bietet der Hauptbereich die folgenden Registerkarten (modellabhängige Registerkarten füllen sich erst nach der Anpassung):

| Registerkarte       | Inhalt                                                                                                                                                            |
|---------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data                | Interaktive DataTable der importierten Daten. In den Modi Appraisal und Market aufgeteilt in Tabellen für Bewertungsobjekt und Vergleichsverkäufe.                |
| Equation            | Vollständige GAM-Formel in MathJax-Darstellung, dazu eine Tabelle der Definitionen der glatten Funktionen.                                                        |
| Summary             | Kennzahlen ( $R^2$ , $CV\text{-}R^2$ , erklärte Devianz, AIC, n, Anzahl der Glättungsterme) sowie Tabellen der glatten und parametrischen Terme.                  |
| Variable Importance | Waagerechtes Balkendiagramm der F-Statistiken (glatt) und $ t $ -Werte (parametrisch), dazu eine Rangtabelle.                                                     |
| Contribution        | Interaktive Plotly-Kurven der glatten partiellen Effekte mit Konfidenzbändern und Markierungen der earth-Knoten.                                                  |
| Correlation         | Heatmap der Korrelationen numerischer Prädiktoren (bereits vor der Anpassung verfügbar).                                                                          |
| Diagnostics         | Residuen gegen angepasste Werte, Q-Q-Diagramm, Histogramm, Zielgröße gegen angepasste Werte (über <b>gratia</b> ). Streudiagramm beobachtet gegen prognostiziert. |
| RCA Adjustments     | Histogramme der RCA-Analyse (Modi Appraisal/Market, nach der Berechnung).                                                                                         |
| Anova               | Kombinierte ANOVA-Tabelle für parametrische und glatte Terme.                                                                                                     |
| MgcV Output         | Rohtext: Laufzeiten, Formel, Modellzusammenfassung, Familie, Concurvity.                                                                                          |
| Sign Check          | Prüfung der Richtungskonsistenz zwischen earth und GAM (bei importiertem earth-Ergebnis).                                                                         |
| Concurvity          | Gesamt- und paarweise Concurvity-Matrizen mit Farbcodierung.                                                                                                      |
| Basis Check         | Tests der Angemessenheit der Basisdimension über <b>gam.check()</b> .                                                                                             |

### 3.5 Dauerhafte Speicherung der Einstellungen

mgcvUI speichert Ihre Konfiguration automatisch in einer SQLite-Datenbank, indiziert nach dem Namen der Eingabedatei. Wenn Sie dieselbe Datei erneut laden, werden alle Einstellungen wiederhergestellt: Zielauswahl, Prädiktor-Kontrollkästchen, Datentypen, mgcv-Parameter und Interaktionsmatrix. Der Purpose ist eine Eigenschaft des aktiven Projekts und keine globale Einstellung; die zuletzt verwendete Eingabedatei wird je Projekt gespeichert und beim erneuten Öffnen des Projekts automatisch geladen.

### 3.6 Dark Mode

Klicken Sie auf das Mond-/Sonnensymbol oben rechts, um zwischen den Themes Nord Light und Nord Dark zu wechseln. Die Theme-Einstellung wird im localStorage gespeichert und bleibt über Sitzungen hinweg erhalten. Alle Oberflächenelemente passen sich dem gewählten Theme an.

### 3.7 Locale & Regional Settings (Gebietsschema und regionale Einstellungen)

mgcvUI unterstützt internationale Konventionen für Zahlen- und CSV-Formate über ein länderbasiertes Locale-System. Das **Settings**-Menü in der Titelleiste bietet Auswahlfelder für **Country** (Land) und **Paper** (Papierformat) für über 30 unterstützte Länder. Jede Voreinstellung legt fest:

- **CSV-Trennzeichen** — Komma (,) für USA/Großbritannien/Japan oder Semikolon (;) für den größten Teil Europas.
- **Dezimaltrennzeichen** — Punkt (.) oder Komma (,).

- **Tausendertrennzeichen** — Komma (USA/Großbritannien/Japan), Punkt (Deutschland/Italien/Spanien), Leerzeichen (Finnland/Frankreich/Polen/Baltikum) oder Apostroph (Schweiz).
- **Papierformat** — Letter (USA/Kanada/Mexiko) oder A4 (überall sonst).

Klicken Sie auf **Save**, um Ihre Locale-Einstellungen global zu speichern. Der Abschnitt Import Data verfügt zudem über ein dateibezogenes Auswahlfeld **Locale** für einmalige Importe mit abweichenden Konventionen. Dasselbe Settings-Menü enthält das Feld **regProj root folder**, mit dem der gemeinsame Projektbaum von mgcvUI, earthUI und glmnetUI verlegt werden kann.

***Hinweis:** Die Zahlenformatierung auf Diagrammachsen und Steigungsbeschriftungen passt sich automatisch an das eingestellte Locale an. Das deutsche Locale verwendet Punkte als Tausendertrennzeichen (200.000), das finnische Leerzeichen (200 000), das schweizerische Apostrophe (200'000). Währungssymbole werden nicht angezeigt — mgcvUI ist währungsneutral.*

## 4 Der Appraisal-Modus

Ist der Purpose des aktiven Projekts **For Appraisal**, richtet sich mgcvUI auf die Bewertung eines einzelnen Objekts ein. Alle in Kapitel 3 beschriebenen Funktionen bleiben verfügbar; dieses Kapitel behandelt nur die wertermittlungsspezifischen Ergänzungen.

### 4.1 Behandlung der Bewertungsobjektzeile

Im Wertermittlungsmodus ist **Zeile 1 Ihres Datensatzes das Bewertungsobjekt**, alle übrigen Zeilen sind Vergleichsverkäufe. Ihre Eingabedatei muss entsprechend aufgebaut sein (siehe Kapitel 2). Der Verkaufspreis des Bewertungsobjekts kann leer bleiben oder einen beliebigen Wert enthalten — mgcvUI behandelt ihn bei der Anpassung automatisch als NA.

Nach der Anpassung erzeugt das Modell Prognosen für die Zeile des Bewertungsobjekts; sie bilden die Grundlage des RCA-Anpassungsablaufs.

### 4.2 Effective Date & Sale Age

In den Modi Appraisal und Market erscheint im Abschnitt Variable Configuration ein Feld **Effective Date** (Wertermittlungstichtag), das mit dem heutigen Datum vorbelegt ist. Kennzeichnen Sie eine Spalte im Special-Auswahlfeld als **contract\_date**, berechnet mgcvUI eine Spalte **sale\_age** — die Anzahl ganzer Tage zwischen dem Vertragsdatum des jeweiligen Verkaufs und dem Stichtag. Diese Spalte wird als Prädiktor ergänzt.

Ändert sich das Effective Date, wird **sale\_age** automatisch neu berechnet.

### 4.3 Besondere Spaltenkennzeichnungen

In den Modi Appraisal und Market erscheint in der Tabelle Variable Configuration für jeden Prädiktor ein Auswahlfeld **Special**. Die vollständige Referenz der Special Types und ihrer Wirkungen finden Sie in Kapitel 6.

### 4.4 Überblick über die RCA-Anpassungen

Die Schaltfläche **Calculate RCA Adjustments & Download** (in den Modi Appraisal und Market nach der Anpassung sichtbar) berechnet marktabgeleitete Anpassungen für jedes Vergleichsobjekt im Verhältnis zum Bewertungsobjekt. Der vollständige RCA-Ablauf ist in Kapitel 11 beschrieben.

***Hinweis:** Der dem Bewertungsobjekt zugewiesene CQA-Wert steuert, welcher Anteil der Residualverteilung dem Bewertungsobjekt zugerechnet wird. Ein Wert von 5.00 platziert das Bewertungsobjekt am Median der Vergleichsobjekte.*

## 5 Der Modus Market Area Analysis

Ist der Purpose des aktiven Projekts **Market Area Analysis**, stellt mgcvUI dieselben immobilien-spezifischen Funktionen bereit wie im Wertermittlungsmodus (Special-Spalten, Sale Age), ist jedoch auf die Analyse einer Objektgruppe statt auf die Bewertung eines einzelnen Objekts ausgerichtet.

### 5.1 Unterschiede zum Appraisal-Modus

- **Flexible Behandlung der Bewertungsobjektzeile** — im Market-Modus wird Zeile 1 nicht automatisch als Bewertungsobjekt behandelt.
- **Kein Sales Grid** — der Schritt „Generate Sales Grid & Download“ steht nicht zur Verfügung.

### 5.2 Wann der Market-Modus sinnvoll ist

Der Modus Market Area Analysis eignet sich, wenn Sie:

- ein GAM-Modell für eine Nachbarschaft oder ein Marktgebiet aufbauen, um nichtlineare Werttreiber zu verstehen
- analysieren, wie Variablen wie Wohnfläche, Alter, Grundstücksgröße und Lage die Verkaufspreise in einer Objektgruppe beeinflussen
- Belege für eine Marktkonditionenanalyse oder eine Abgrenzung der Nachbarschaft vorbereiten
- mit einem Datensatz arbeiten, für den Sie die Special-Spalten-Funktionen (Sale Age) benötigen, den Sales-Grid-Ablauf jedoch nicht

**Hinweis:** Der Market Mode eignet sich auch für allgemeine Immobilienregressionen, bei denen die Special-Spalten-Funktionen benötigt werden, der Ablauf zur Bewertung eines einzelnen Objekts jedoch nicht.

## 6 Variablenauswahl

In Abschnitt 4 der Seitenleiste — **Variable Configuration** — legen Sie fest, welche Spalten in das Modell eingehen und wie sie behandelt werden.

### 6.1 Zielvariable

Das Auswahlfeld **Target (response) variable** am Kopf von Abschnitt 4 listet jede numerische Spalte Ihres Datensatzes auf. Wählen Sie eine Spalte als Zielvariable (etwa den Verkaufspreis). Die Zielspalte wird automatisch aus der Prädiktorliste ausgeschlossen.

### 6.2 Transformation der Zielgröße

Unterhalb der Zielauswahl bietet ein Auswahlfeld **Response Transform** drei Möglichkeiten:

- **None (raw)** — Die Zielgröße wird unverändert verwendet. Angebracht, wenn der Zusammenhang zwischen Prädiktoren und Zielgröße auf der ursprünglichen Skala näherungsweise additiv ist.
- **Log (natural)** — Vor der Anpassung wird der natürliche Logarithmus angewandt. Nach der Rücktransformation werden die Anpassungen dadurch proportional (%) statt absolut (\$). Besonders nützlich bei Verkaufspreisen, wo eine Anpassung von 10.000 \$ bei unterschiedlichen Preisniveaus sehr Unterschiedliches bedeutet.
- **Log10** — Es wird der Zehnerlogarithmus angewandt.

***Hinweis:** Die Log-Transformation macht Zeitanpassungen proportional (%) statt absolut (\$). Bei Immobilienmodellen über eine breite Preisspanne führen Log-Transformationen häufig zu besser verteilten Residuen.*

Ist eine Log-Transformation gewählt, werden Werte  $\leq 0$  in der Zielgröße automatisch herausgefiltert. Sämtliche Ausgaben (Prognosen, Beiträge, Residuen) werden automatisch auf die ursprüngliche Skala zurücktransformiert.

### 6.3 Die Prädiktortabelle

Unterhalb der Zielauswahl listet eine scrollbare Tabelle alle übrigen Spalten auf. Die Kontrollkästchenspalten verwenden aus Platzgründen gedrehte vertikale Kopfzeilen. Die Spaltenreihenfolge lautet:

| Spalte   | Beschreibung                                                                                                                       |
|----------|------------------------------------------------------------------------------------------------------------------------------------|
| Variable | Spaltenname (Festbreitenschrift, bei Bedarf mit Auslassungspunkten gekürzt). Umrandete Zelle.                                      |
| Type     | Auswahlfeld für den Datentyp: numeric, integer, character, factor, logical, Date, POSIXct. Umrandetes Auswahlfeld.                 |
| Include  | Kontrollkästchen — diese Spalte als Prädiktor in das Modell aufnehmen. Vertikale Kopfzeile.                                        |
| Factor   | Kontrollkästchen — als kategorialen Faktor behandeln (erzeugt je Stufe einen eigenen Koeffizienten). Vertikale Kopfzeile.          |
| Linear   | Kontrollkästchen — linearen Zusammenhang erzwingen (kein Glättungsterm, nur $\beta x$ ). Vertikale Kopfzeile.                      |
| Special  | Auswahlfeld (nur Appraisal/Market) — siehe Referenz der Special-Spaltentypen weiter unten. Umrandetes Auswahlfeld.                 |
| NAs      | Anzahl und Anteil fehlender Werte, farblich codiert: rot ( $\geq 50\%$ ), orange ( $\geq 20\%$ ), grau ( $> 0$ ). Umrandete Zelle. |

**Factor vs. Smooth vs. Linear:** Standardmäßig erhalten einbezogene numerische Variablen einen Glättungsterm  $f(x)$ . Das Setzen von **Factor** erzeugt einen kategorialen Term (einen Koeffizienten je Stufe). Das Setzen von **Linear** erzeugt statt eines Glättungsterms einen einfachen linearen Term  $\beta x$ . Variablen, die sowohl als Factor als auch als Linear markiert sind, werden als Factor behandelt.

## 6.4 Erkennung und Überschreiben von Datentypen

mgcvUI erkennt die Datentypen beim Import automatisch. Numerische, ganzzahlige, logische, Faktor- und Datumsspalten werden erkannt. Sie können jede Erkennung über das Auswahlfeld **Type** überschreiben. Eine Typänderung wirkt sich darauf aus, wie die Spalte im Modell behandelt wird.

## 6.5 Referenz der Special-Spaltentypen

In den Modi Appraisal und Market bietet das Auswahlfeld **Special** die folgenden Optionen:

### Gewichtung:

- **weight** — Spalte mit Beobachtungsgewichten (nur eine zulässig; Zeilen mit Gewicht = 0 werden von der Anpassung ausgeschlossen)

### Datums- und Zeittypen:

- **contract\_date** — Löst die automatische Berechnung von **sale\_age** ab dem Effective Date aus
- **sale\_age** — Kennzeichnet eine bereits berechnete Sale-Age-Spalte (Tage ab dem Stichtag), die unverändert verwendet wird
- **listing\_date** — Wird als Rückfalloption für die Berechnung der Vermarktungsdauer verwendet
- **dom** — Kennzeichnet die Spalte mit der Vermarktungsdauer (Days on Market)

### Datensatztyp:

- **sale\_type** — Kennzeichnet die Spalte des Datensatztyps, die das Bewertungsobjekt von den Vergleichsverkäufen unterscheidet

### Monetäre Typen:

- **concessions** — Kennzeichnet Verkaufszugeständnisse (Verkäuferzuschüsse, Käuferanreize usw.)

### Größen- und Lagetypen:

- **latitude** — Werte werden automatisch auf 3 Nachkommastellen gerundet
- **longitude** — Gleiche Rundung wie bei latitude
- **area** — Kennung des Marktgebiets bzw. der Nachbarschaft
- **living\_area** — Ermöglicht die Berechnung von Residuen je Flächeneinheit (**residual\_sf** und **cqa\_sf**)
- **lot\_size** — Spalte für die Grundstücksgröße
- **site\_dimensions** — Wird mit der Grundstücksgröße gruppiert

### Alterstypen:

- **actual\_age** — Spalte für das tatsächliche Objektalter
- **effective\_age** — Effektives Objektalter

### Anzeigetypen:

- **display\_only** — Die Spalte wird in die Excel-Exporte aufgenommen, aber vollständig von der Modellanpassung ausgeschlossen. Geeignet für Adressfelder, MLS-Nummern, Flurstückskennungen oder sonstige Referenzdaten.

***Hinweis:*** Als ***display\_only*** gekennzeichnete Spalten verbleiben in Ihrem Datensatz und erscheinen im Excel-Export, werden jedoch vollständig aus der Prädiktortabelle ausgeschlossen. Diese Kennzeichnung eignet sich für ID-Spalten, Adressen oder sonstige Referenzfelder.

## 7 Parameterwahl

Abschnitt 5 der Seitenleiste — **Mgcv Call Parameters** — eröffnet den Zugang zu sämtlichen Konfigurationsoptionen des GAM. Jeder Parameter besitzt ein blaues Hilfesymbol (?) mit einer Erläuterung als Kurzinfo.

### 7.1 Parameter-Voreinstellungen

Ein Auswahlfeld am Kopf bietet zwei Voreinstellungen mit sinnvollen Standardwerten für gängige Arbeitsabläufe:

| Voreinstellung                   | Basis                    | Einstellungen                             |
|----------------------------------|--------------------------|-------------------------------------------|
| Standalone<br>(Erkundung)        | tp (Thin Plate)          | k = auto (10), gamma = 1.2, select = TRUE |
| Earth Pipeline<br>(Verfeinerung) | cr (kubische Regression) | k = auto (10), gamma = 1.4, select = TRUE |

Die Voreinstellung **Earth Pipeline** wird automatisch gewählt, wenn earthUI-Knoten importiert werden. Die Basis des kubischen Regressionssplines (**cr**) ist für die Einbindung der earth-Knoten erforderlich, weil sie die Angabe exakter Knotenpositionen erlaubt.

### 7.2 Family (Verteilungsfamilie)

Wählen Sie die Verteilungsfamilie für Ihre Zielvariable:

| Familie                 | Anwendungsfall                                                             |
|-------------------------|----------------------------------------------------------------------------|
| <b>gaussian</b>         | Stetige Zielgrößen (z. B. Verkaufspreis). Am gebräuchlichsten.             |
| <b>Gamma</b>            | Positive stetige Zielgrößen mit einer Varianz proportional zum Mittelwert. |
| <b>poisson</b>          | Zählraten (z. B. Anzahl der Verkäufe).                                     |
| <b>binomial</b>         | Binäre Ergebnisse.                                                         |
| <b>inverse.gaussian</b> | Stark rechtsschiefe positive Daten.                                        |

### 7.3 Method (Schätzverfahren)

Das Verfahren zur Schätzung des Glättungsparameters:

| Verfahren      | Beschreibung                                                                            |
|----------------|-----------------------------------------------------------------------------------------|
| <b>REML</b>    | Restricted Maximum Likelihood (Vorgabe, empfohlen). Am robustesten gegen Überanpassung. |
| <b>GCV.Cp</b>  | Generalisierte Kreuzvalidierung. Neigt dazu, etwas komplexere Modelle zu wählen.        |
| <b>ML</b>      | Maximum Likelihood. Ähnlich wie REML, kann aber zu wenig glätten.                       |
| <b>P-REML</b>  | Pearson-Variante von REML.                                                              |
| <b>P-ML</b>    | Pearson-Variante von ML.                                                                |
| <b>GACV.Cp</b> | Generalisierte approximative Kreuzvalidierung.                                          |

## 7.4 Gamma (Glättungsmultiplikator)

Der Parameter gamma multipliziert die effektiven Freiheitsgrade im Glättungskriterium und begünstigt damit glattere (weniger wellige) Anpassungen. Der Vorgabewert ist 1.2.

| Wert    | Wirkung                                                                   |
|---------|---------------------------------------------------------------------------|
| 1.0     | Standardglättung                                                          |
| 1.2–1.4 | Etwas glatter als die Vorgabe (empfohlen)                                 |
| 2.0+    | Deutlich glatter — geeignet für verrauschte Daten oder kleine Stichproben |

Höhere Gamma-Werte schützen vor Überanpassung, indem sie Komplexität stärker bestrafen. Die Voreinstellung Earth Pipeline verwendet 1.4, um beim Verfeinern der earth-Knoten zusätzlich zu glätten.

## 7.5 Kreuzvalidierung

Ist das Kästchen gesetzt (Vorgabe), führt mgcvUI nach der Anpassung eine 10-fache Kreuzvalidierung durch und berechnet ein CV-Bestimmtheitsmaß. Dieses liefert eine ehrliche Schätzung der Prognosegüte außerhalb der Stichprobe.

## 7.6 Select (Schrumpfung)

Ist das Kästchen gesetzt (Vorgabe), erhält jeder Glättungsterm einen zusätzlichen Strafterm, der ihn auf null schrumpfen kann. Damit wird eine automatische Variablenselektion möglich — unwichtige Glättungsterme werden faktisch aus dem Modell entfernt. Umgesetzt über `select = TRUE` in `gam()`.

## 7.7 Default Basis (Standardbasis)

Der Typ der Spline-Basisfunktion für Glättungsterme:

| Basis     | Vollständige Bezeichnung     | Beschreibung                                                                                |
|-----------|------------------------------|---------------------------------------------------------------------------------------------|
| <b>tp</b> | Thin-Plate-Regressionsspline | Vorgabe. Isotrop (keine Knotensetzung erforderlich). Gute Allzweckwahl.                     |
| <b>cr</b> | Kubischer Regressionsspline  | Erlaubt die explizite Setzung von Knoten. Für die Einbindung der earth-Knoten erforderlich. |
| <b>ps</b> | P-Spline                     | Gleichmäßig verteilte B-Spline-Basis mit Differenzen-Strafterm. Rechnerisch effizient.      |
| <b>bs</b> | B-Spline                     | Flexible B-Spline-Basis.                                                                    |

## 7.8 Default k (Basisdimension)

Die Basisdimension  $k$  steuert die maximale Welligkeit jedes Glättungsterms. Der Wert 0 (Vorgabe) bedeutet „automatisch“ — mgcvUI verwendet  $k = 10$  oder die Anzahl der earth-Knoten, je nachdem, was angemessen ist.

- **Kleines  $k$  (3–5):** Sehr glatte, nahezu lineare Kurven.
- **Mittleres  $k$  (8–15):** Mäßige Flexibilität (meist ausreichend).
- **Großes  $k$  (20+):** Sehr flexibel — Gefahr der Überanpassung ohne starke Regularisierung.

**Hinweis:** Die effektiven Freiheitsgrade (EDF) in der Registerkarte Summary zeigen an, wie viele Freiheitsgrade jeder Glättungsterm tatsächlich nutzt. Liegt EDF nahe bei  $k - 1$ , ist die Basis möglicherweise zu klein — erhöhen Sie  $k$ . Ist EDF deutlich kleiner als  $k$ , ist der Glättungsterm gut penalisiert.

## 7.9 Tensorotyp

Für Interaktionen zwischen stetigen Variablen stehen zwei Tensorprodukt-Typen zur Verfügung:

| Typ       | Funktion          | Beschreibung                                                                                                                                                                             |
|-----------|-------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>ti</b> | <b>ti(x1, x2)</b> | Tensor-Interaktion — modelliert nur den Interaktionseffekt; die Haupteffekte werden getrennt durch univariate <b>s()</b> -Terme abgebildet. Wegen der Interpretierbarkeit zu bevorzugen. |
| <b>te</b> | <b>te(x1, x2)</b> | Tensorprodukt — modelliert den gesamten gemeinsamen Effekt einschließlich der Haupteffekte. Für RCA-Anpassungen schwerer zu zerlegen.                                                    |

## 7.10 Zugelassene Interaktionen

Ein aufklappbarer Abschnitt **Allowed Interactions** zeigt eine obere Dreiecksmatrix aus Kontrollkästchen für alle einbezogenen glatten (nicht linearen, nicht faktoriellen) Prädiktoren. Jedes Kästchen aktiviert einen Tensorprodukt-Glättungsterm zwischen dem zugehörigen Variablenpaar.

- **Allow All / Clear All** — Schaltflächen zum Umschalten aller Kästchen.
- **Klick auf einen Variablennamen** — schaltet alle Interaktionen dieser Variable um.
- **Rechtsklick auf einen Variablennamen** — „Block from main effect“ — markiert die Variable zur ausschließlichen Verwendung in Interaktionen (kein univariater Glättungsterm).

Werden earthUI-Knoten importiert, sind die von earth erkannten Interaktionen vorab aktiviert und gesperrt (gelb hinterlegt). Hat das earth-Modell  $\text{degree} = 1$  verwendet (keine Interaktionen), erscheint ein entsprechender Hinweis.

### Warnung

Das Aktivieren von Tensor-Interaktionen erhöht die Modellkomplexität erheblich. Setzen Sie Interaktionen ein, wenn Fachwissen oder von earth erkannte Belege dafür sprechen. Prüfen Sie, ob die resultierenden Anpassungen plausibel und erklärbar sind.

## 7.11 Faktor-Glättungsterm-Interaktionen

Sind sowohl Faktor- als auch Glättungsvariablen einbezogen, erscheint eine gesonderte Matrix **Factor-by-Smooth Interactions**. Jedes Kästchen erzeugt einen Term **s(x, by = factor)** — eine eigene glatte Kurve je Faktorstufe.

## 7.12 Erweiterte Parameter

Ein aufklappbarer Abschnitt „Advanced“ macht weitere Einstellungen zugänglich:

- **Optimizer** — Algorithmus zur Schätzung des Glättungsparameters: `outer/newton` (Vorgabe), `outer/bfgs` oder `efs` (Extended Fellner-Schall).
- **Scale** — Skalenparameter. 0 (Vorgabe) = aus den Daten schätzen. Für Modelle mit fester Skala auf einen bekannten Wert setzen.
- **Discrete covariate method** — Schnelle diskretisierte Anpassung für große Datensätze (`discrete = TRUE`).
- **Threads** — Anzahl paralleler Threads (1–32). Nur wirksam, wenn `discrete = TRUE`.

## 8 Anpassen des Modells

### 8.1 Die Schaltfläche Fit

Abschnitt 6 der Seitenleiste enthält die Schaltfläche **Fit MgcV GAM Model**. Ein Klick darauf führt das Modell mit Ihrer aktuellen Konfiguration aus.

### 8.2 Was während der Anpassung geschieht

Wenn Sie auf Fit klicken, geschieht in mgcvUI Folgendes:

1. **Prüfung der Konfiguration** — es wird geprüft, ob mindestens ein Prädiktor und mindestens 10 vollständige Zeilen vorliegen. Sind weniger als 50 % der Zeilen vollständig, erscheint eine Warnung (mit Angabe der Spalten, die die meisten NAs verursachen).
2. **Aufbereitung der Daten** — Beschränkung auf die für das Modell relevanten Spalten, Entfernen von Zeilen mit NAs, Anwenden von Gewichten (sofern gekennzeichnet), Ausschluss der Bewertungsobjektzeile (im Wertermittlungsmodus), Anwenden der Transformation der Zielgröße.
3. **Aufbau der GAM-Formel** — es werden Glättungsterme `s(x, bs=cr", k=10)` für numerische Prädiktoren, Faktorterm für kategoriale Variablen, lineare Terme für als Linear markierte Variablen und Tensorprodukte für die gewählten Interaktionen gebildet. Earth-Knoten werden als explizite Knotenpositionen in Terme der `cr`-Basis eingebettet.
4. **Behandlung von Sonderfällen** — konstante Variablen werden verworfen, binäre Prädiktoren in Faktoren umgewandelt,  $k$  auf die Anzahl der eindeutigen Werte minus 1 begrenzt und die earth-Knoten um Datenquantile ergänzt, wenn mehr Knoten benötigt werden.
5. **Anpassung des GAM** — `mgcv::gam()` wird mit der gebildeten Formel sowie mit `family`, `method`, `gamma`, `select` und den weiteren Parametern aufgerufen.
6. **Kreuzvalidierung** (sofern aktiviert) — 10-fache Kreuzvalidierung zur Berechnung des Bestimmtheitsmaßes außerhalb der Stichprobe.
7. **Aktualisierung aller Ergebnis-Registerkarten** — Summary, Equation, Contribution-Diagramme, Diagnostics und alle übrigen Registerkarten werden gefüllt.

### 8.3 Anpassung im Hintergrund & Fortschritt

Steht das Paket `callr` zur Verfügung, läuft die Modellanpassung in einem Hintergrundprozess, sodass die Anwendung bedienbar bleibt. Eine dunkel gehaltene, terminalartige Fortschrittsanzeige zeigt:

- einen **Timer** mit den verstrichenen Sekunden bzw. Minuten.
- ein **farbcodiertes Trace-Protokoll**: blau für Status, gelb für den Fortschritt der CV-Faltungen, rot für Fehler, grün für Ergebnisse.
- eine Schaltfläche **Abort**, um den Anpassungsprozess bei Bedarf abubrechen.
- eine Schaltfläche **Close**, die nach Abschluss erscheint.

Ist `callr` nicht verfügbar, läuft die Anpassung synchron (die Anwendung ist bis zum Abschluss blockiert).

Nach erfolgreicher Anpassung erscheint ein weißes Häkchen auf der Schaltfläche Fit, und eine Statuszeile zeigt: „R-sq = X, CV R-sq = X, Dev = X%, AIC = X, n = X“.

## 9 Ergebnis-Registerkarten

### 9.1 Data

Zeigt die importierten Daten als interaktive DataTable mit waagerechtem Bildlauf und 15 Zeilen je Seite. In den Modi Appraisal und Market ist die Registerkarte in eine Tabelle **Subject Property** (Zeile 1) und eine Tabelle **Comparable Sales** (die übrigen Zeilen) unterteilt; im General-Modus wird eine einzige Tabelle angezeigt.

Jede Zelle bleibt einzeilig und wird auf die Spaltenbreite gekürzt, damit breite Freitextfelder (etwa Objektbemerkungen) eine Zeile nicht über die Seite ziehen. Ein einfacher Klick wählt eine Zeile aus; ein **Doppelklick auf eine beliebige Zelle** öffnet ein Fenster mit ihrem vollständigen, ungekürzten Inhalt.

### 9.2 Equation

Zeigt das angepasste GAM-Modell in zwei Teilen:

**Modellgleichung** (in MathJax dargestellt):

- Die linke Seite zeigt die Zielgröße mit der zugehörigen Transformation (log, log10) und der Linkfunktion.
- Die rechte Seite zeigt den Wert des Achsenabschnitts, die Glättungsterme als  $f_i(\text{Variable})$ , die nach Variablen gruppierten Faktorterm mit Anzahl der Stufen sowie die linearen Terme mit Koeffizientenwerten.
- Familie und Verfahren werden oberhalb der Gleichung angezeigt.

Tabelle **Smooth Function Definitions**:

- Formelausdruck des Terms (z. B. `s(living_sqft, bs=cr", k=8)`)
- Variablenname
- Typ (s, te, ti oder linear)
- Typ der Basisfunktion
- Wert von  $k$
- Positionen der earth-Knoten (sofern importiert)

### 9.3 Summary

Sechs Kennzahlenkarten am oberen Rand:

- **R-squared** — Anteil der erklärten Devianz
- **CV R-squared** — kreuzvalidiertes Bestimmtheitsmaß (bei aktivierter Kreuzvalidierung)
- **Dev. Explained %** — Prozentsatz der erklärten Devianz
- **AIC** — Akaike-Informationskriterium
- **n** — Anzahl der verwendeten Beobachtungen
- **Smooths** — Anzahl der Glättungsterme

Unterhalb der Karten:

- **Tabelle Smooth Terms** — Term, EDF (effektive Freiheitsgrade), Ref.df, F-Statistik, p-Wert
- **Tabelle Parametric Terms** — Term, Schätzwert, Standardfehler, t-Wert, p-Wert

### 9.4 Variable Importance

Ein waagerechtes Balkendiagramm zeigt die relative Bedeutung jedes Modellterms:

- **Glättungsterme**: nach F-Statistik geordnet (blaue Balken)
- **Parametrische Terme**: nach  $|t|$ -Wert geordnet (grüne Balken)

Unter dem Diagramm zeigt eine DataTable Term, Typ, EDF, Statistik und p-Wert, absteigend nach Statistik sortiert.

## 9.5 Beitragsdiagramme

Jeder Modellterm, der zum Prognosewert beiträgt, besitzt ein interaktives **Plotly**-Beitragsdiagramm, das in einer umrandeten Karte dargestellt wird. Die Registerkarte zeigt Diagramme für alle Termtypen:

### 9.5.1 Univariate Glättungsterme — $s(x)$

- **Blaue Linie** mit 95 %-Konfidenzband
- **Graue Streupunkte** für die partiellen Residuen (bei nicht log-transformierten Modellen)
- **Rote gestrichelte Senkrechte** an den Positionen der earth-Knoten (sofern earth importiert wurde)
- **Kurzinfo beim Überfahren**: Variablenwert, Beitrag (\$), 95 %-Konfidenzintervall, Rate (Steigung je Einheit)

Bei **log-transformierten Modellen** wird die y-Achse mit der Formel  $\bar{y} \times (e^{f(x)} - 1)$  in Geldbeträge zurücktransformiert, sodass die Kurven unmittelbar in Geldeinheiten interpretierbar sind.

Für die Variablen **latitude/longitude** wird die Steigung je 0,001 Grad statt je Grad angegeben, da dies der aussagekräftigere geografische Zuwachs ist.

### 9.5.2 Faktor-Glättungsterme — $s(x, \text{by} = \text{Faktor})$

Eine farbige Linie je Faktorstufe, in einem gemeinsamen Diagramm überlagert. Nützlich, um zu erkennen, wie der Effekt eines Prädiktors zwischen Gruppen variiert (etwa zwischen verschiedenen Nachbarschaften).

### 9.5.3 Interaktionsterme — $ti(x_1, x_2)$ und $te(x_1, x_2)$

Eine **Heatmap** zeigt die zweidimensionale Beitragsfläche. Farbskala: rot (negativ) → weiß (null) → blau (positiv). Beim Überfahren werden beide Variablenwerte und der Beitragsbetrag angezeigt.

### 9.5.4 Parametrische Terme — linear und faktoriell

- **Numerische lineare Terme**: Streudiagramm der Datenpunkte mit überlagerter Anpassungsgeraden
- **Faktorterm**: Balkendiagramm des mittleren Beitrags je Faktorstufe

## 9.6 Correlation

Eine Heatmap-Matrix der Pearson-Korrelationen zwischen allen numerischen Prädiktoren. Unmittelbar nach dem Datenimport verfügbar — eine Modellanpassung ist nicht erforderlich.

- Farbverlauf: blau (−1) → weiß (0) → rot (+1)
- Korrelationswerte werden in jede Zelle geschrieben
- Nützlich, um bereits vor der Anpassung Concurvity (das glatte Analogon zur Multikollinearität) zu erkennen

## 9.7 Diagnostics

Zwei Gruppen diagnostischer Darstellungen:

**Diagnosetafel** (über `gratia::appraise`): Vier Diagramme — Residuen gegen angepasste Werte, Q-Q-Diagramm, Histogramm der Residuen sowie Zielgröße gegen angepasste Werte.

Streudiagramm **Actual vs Predicted**: beobachtete gegen prognostizierte Werte mit einer gestrichelten 45-Grad-Referenzlinie. Bei log-transformierten Modellen werden die Werte auf die ursprüngliche Skala zurücktransformiert.

## 9.8 RCA Adjustments

Drei Histogramme, die nach der RCA-Berechnung angezeigt werden (nur in den Modi Appraisal/Market):

- **Residual Adjustment %** — Verteilung der Residualanpassungen als Prozentsatz des Verkaufspreises
- **Net Adjustment %** — Verteilung der Nettoanpassungen
- **Gross Adjustment %** — Verteilung der Bruttoanpassungen

Jedes Histogramm nennt in der Unterzeile Mittelwert, Median und Standardabweichung und zeigt gestrichelte Referenzlinien für Mittelwert und Median.

## 9.9 Anova

Eine kombinierte ANOVA-Tabelle, die parametrische und glatte Terme aus der GAM-Zusammenfassung zusammenführt. Sie zeigt Term, Typ (parametrisch oder glatt) und alle zugehörigen Statistiken.

## 9.10 MgcV Output

Rohe Textausgabe mit:

- Laufzeit der Anpassung
- der GAM-Formel
- der Druckausgabe des Modells
- der vollständigen Ausgabe von `summary(model)`
- Familie und Linkfunktion
- Gesamtmaßen der Concurvity

## 9.11 Sign Check

Werden earthUI-Knoten importiert, vergleicht diese Registerkarte die Wirkungsrichtung jeder Variablen im earth-Modell mit ihrer Richtung im GAM:

- **earth\_direction**: steigend, fallend oder gemischt (auf Basis der Vorzeichen der Hinge-Funktionen)
- **gam\_direction**: steigend, fallend oder gemischt (aus der Ableitung des Glättungsterms auf einem Gitter mit 100 Punkten geschätzt)
- **consistent**: TRUE/FALSE
- **warning**: Beschreibung etwaiger Abweichungen

Eine Statuszeile zeigt „All directions consistent“ oder die Anzahl der Variablen mit Abweichungen an.

### 9.12 Concurvity

Concurvity ist das glatte Analogon zur Multikollinearität. Diese Registerkarte zeigt:

- **Tabelle Overall Concurvity:** Zeilen für die Maße worst/observed/estimate, Spalten für jeden Glättungsterm. Farblich codiert: weiß ( $<0,5$ ), gelb ( $0,5-0,8$ ), rot ( $>0,8$ ).
- **Tabelle Pairwise Concurvity:** paarweise Matrix der ungünstigsten Fälle mit derselben Farbcodierung.

Werte oberhalb von etwa 0,8 in der Zeile „worst“ deuten darauf hin, dass sich ein Glättungsterm gut durch die übrigen Glättungsterme annähern lässt, was auf Redundanz hinweist.

### 9.13 Basis Check

Ausgabe von `mgcv::gam.check()`: Tests der Angemessenheit der Basisdimension für jeden Glättungsterm. Ist das  $k$  eines Glättungsterms zu klein, weist der Test dies mit einem niedrigen p-Wert aus und legt eine Erhöhung von  $k$  nahe.

## 10 Herunterladen der Daten

Nach der Anpassung exportiert die Schaltfläche **Download Output (Excel)** (Abschnitt 7 der Seitenleiste) eine Excel-Datei mit Prognosen und Diagnosewerten in den mgcv-Ausgabeordner des aktiven Projekts.

### 10.1 Ausgabespalten

| Spalte             | Beschreibung                                                                                                  |
|--------------------|---------------------------------------------------------------------------------------------------------------|
| basis              | Rücktransformierter Wert des Achsenabschnitts (für alle Zeilen gleich)                                        |
| <var>_contribution | Beitrag des jeweiligen Glättungsterms zur Prognose. Bei Log-Modellen proportional auf Geldbeträge aufgeteilt. |
| est_<target>       | Modellprognose (z. B. <code>est_sale_price</code> ), gegebenenfalls von der Log-Skala zurücktransformiert     |
| residual           | Beobachtet – prognostiziert                                                                                   |
| calc_residual      | Kontrolle: beobachtet – Summe der Beiträge (rücktransformiert)                                                |
| cqa                | Condition-Quality-Appeal-Wert (CQA, Skala 0–10)                                                               |
| residual_sf        | Residuum / Wohnfläche (sofern <code>living_area</code> gekennzeichnet ist)                                    |
| cqa_sf             | CQA, berechnet aus der Rangfolge über <code>residual_sf</code>                                                |

Im Wertermittlungsmodus sind `residual/cqa/cqa_sf` in Zeile 1 (Bewertungsobjekt) auf NA gesetzt. Ist eine `living_area`-Spalte gekennzeichnet, werden die Zeilen absteigend nach `residual_sf` sortiert.

Nach erfolgreichem Abschluss erscheint ein weißes Häkchen auf der Download-Schaltfläche.

### 10.2 CQA-Werte

CQA stellt das Residuum jeder Zeile auf einer Skala von 0 bis 10 in eine Rangfolge gegenüber allen anderen:

- **Hoher CQA-Wert (etwa 9–10):** deutlich über der Prognose verkauft — vermutlich überdurchschnittlicher Zustand, überdurchschnittliche Qualität oder Attraktivität
- **Niedriger CQA-Wert (etwa 0–1):** deutlich unter der Prognose verkauft — vermutlich schlechterer Zustand oder Notverkauf
- **CQA um 5:** nahe dem Median

## 11 RCA-Berechnungen und Download

Der RCA-Ablauf (Residual Constraint Approach) steht in den Modi **Appraisal** und **Market Area Analysis** nach der Modellanpassung zur Verfügung. Das Bewertungsobjekt muss in Zeile 1 stehen.

### 11.1 Öffnen des RCA-Dialogs

Klicken Sie in Abschnitt 8 der Seitenleiste auf die Schaltfläche **Calculate RCA Adjustments & Download**. Es erscheint ein modaler Dialog mit:

- **Score type:** CQA oder CQA per SF (sofern `living_area` gekennzeichnet ist). Die Variante je Flächeneinheit skaliert das Residuum des Bewertungsobjekts mit der Wohnfläche.
- **Subject CQA Score:** Zulässig sind Werte von 0.00 bis 10.00. Das Feld ist zunächst leer; wurde ein earthUI-Modell importiert, werden der in earthUI verwendete Score-Typ und Wert aus den gemeinsamen Projekteinstellungen übernommen.

Klicken Sie auf **Generate**, um zu berechnen und herunterzuladen. Die Excel-Datei wird in den mgcv-Ausgabeordner des aktiven Projekts geschrieben (`<os>_out_mgcv`).

### 11.2 Interpolation des CQA-Werts

1. Die CQA-Werte und Residuen der Vergleichsobjekte werden sortiert
2. Eine lineare Interpolation (`approx()`) bildet Ihren CQA-Wert auf ein Residuum ab
3. Wert des Bewertungsobjekts = Modellprognose + interpoliertes Residuum
4. Zeilen mit Gewicht 0 werden so behandelt, als hätten sie dasselbe Residuum wie das Bewertungsobjekt

### 11.3 Ausgabespalten

| Spalte                              | Beschreibung                                                                     |
|-------------------------------------|----------------------------------------------------------------------------------|
| <code>subject_value</code>          | Modellprognose + interpoliertes Residuum                                         |
| <code>subject_cqa</code>            | Vom Anwender eingegebener CQA-Wert                                               |
| <code>&lt;var&gt;_adjustment</code> | Beitrag des Bewertungsobjekts – Beitrag des Vergleichsobjekts (in Geldeinheiten) |
| <code>residual_adjustment</code>    | Residuum des Bewertungsobjekts – Residuum des Vergleichsobjekts                  |
| <code>net_adjustments</code>        | Summe aller Anpassungen                                                          |
| <code>gross_adjustments</code>      | Summe der Absolutbeträge der Anpassungen                                         |
| <code>residual_adj_pct</code>       | Residualanpassung in % des Verkaufspreises                                       |
| <code>net_adj_pct</code>            | Nettoanpassungen in % des Verkaufspreises                                        |
| <code>gross_adj_pct</code>          | Bruttoanpassungen in % des Verkaufspreises                                       |
| <code>adjusted_sale_price</code>    | Tatsächlicher Verkaufspreis + Nettoanpassungen                                   |

Nach erfolgreicher Berechnung erscheint ein weißes Häkchen auf der Schaltfläche.

## 12 Vergleichswertraster (Sales Comparison Grid)

Das Vergleichswertraster steht ausschließlich im Modus **Appraisal** zur Verfügung, und zwar nach der Berechnung der RCA-Anpassungen. Es erzeugt eine formatierte Excel-Arbeitsmappe, die sich zur Aufnahme in Wertermittlungsberichte eignet.

### 12.1 Dialog zur Auswahl der Vergleichsobjekte

Ein Klick auf die Schaltfläche **Generate Sales Grid & Download** öffnet einen modalen Dialog, der alle Vergleichsverkäufe auflistet. Die Vergleichsobjekte sind in zwei Gruppen unterteilt:

- **Empfohlene Vergleichsobjekte** (vorab ausgewählt): Bruttoanpassung  $< 25\%$  des Verkaufspreises, sortiert nach Verkaufsalter (jüngste zuerst). Höchstens 30 Vergleichsobjekte.
- **Weitere Vergleichsobjekte** (standardmäßig nicht ausgewählt): die übrigen Vergleichsobjekte, sortiert nach dem Prozentsatz der Bruttoanpassung.

Wählen Sie Ihre Vergleichsobjekte aus und klicken Sie dann auf **Generate Sales Grid**, um die Arbeitsmappe zu erstellen.

**Erforderliche Special Types:** `contract_date` und `living_area` müssen gekennzeichnet sein. **Empfohlene Special Types:** `latitude`, `longitude` und `lot_size` — das Raster funktioniert auch ohne sie, einzelne Felder bleiben dann jedoch leer.

### 12.2 Aufbau der Arbeitsmappe

Die erzeugte Arbeitsmappe enthält mehrere Blätter mit jeweils drei nebeneinander angeordneten Vergleichsobjekten:

- **Kopfzeilen:** Adressen von Bewertungs- und Vergleichsobjekten, Verkaufspreise, Verkaufsdaten und Vermarktungsdauer
- **Gruppierte Zeilen:** Lage (Breitengrad, Längengrad, Gebiet), Grundstücksgröße (Grundstücksfläche, Grundstücksmaße) und Alter (tatsächliches Alter, effektives Alter) mit Gruppenüberschriften und Unterpunkten
- **Zeilen der Modellvariablen:** eine Zeile je Glättungsterm mit dem Beitrag des Bewertungsobjekts, den Beiträgen der Vergleichsobjekte und den Anpassungsbeträgen
- **Zeilen für Restmerkmale:** leere, entsperrte Zellen für Eingaben des Sachverständigen zu Merkmalen, die das Modell nicht erfasst (etwa Zustand, Qualität, Aussicht)
- **Zusammenfassungszeilen:** Nettoanpassung in %, Bruttoanpassung in % und angepasster Verkaufspreis mit aktiven Excel-Formeln

### 12.3 Blattschutz & Zellen für Restmerkmale

Jedes Blatt ist gegen versehentliche Änderungen geschützt; die Zellen für die Werte der Restmerkmale sind jedoch ausdrücklich entsperrt. So können Sachverständige Werte für Merkmale eintragen, die nicht im Modell enthalten sind, während die formelgesteuerten Anpassungsberechnungen erhalten bleiben.

## 13 Herunterladen von Berichten

Die Berichtserstellung erfolgt in einem zweistufigen Quarto-Ablauf, der den einstufigen Bericht früherer Versionen ersetzt.

### 13.1 Schritt 1: Generate Quarto Report

Die Schaltfläche **Generate Quarto Report** (verfügbar, sobald das Modell angepasst ist) schreibt ein in sich geschlossenes `.qmd`-Paket in den `mgcv`-Ausgabeordner des aktiven Projekts. Das Paket enthält den Quarto-Quelltext, vorab gerenderte Diagrammdateien (PNG und PDF), eine Datei `report_data.rds` mit allen Tabellen und Modellmetadaten sowie eine Formatvorlage `reference.docx` für die Word-Ausgabe. Da das Paket in sich geschlossen ist, können Sie den `.qmd`-Quelltext vor dem Rendern bearbeiten oder mit anderen Quarto-Dokumenten kombinieren.

Die Diagramme werden bei der Erzeugung des Pakets vorab gerendert; schlägt ein einzelnes Diagramm fehl, wird ein Platzhalterbild geschrieben, und das übrige Paket wird normal erzeugt.

### 13.2 Schritt 2: Convert Quarto Report

Der Abschnitt **Convert Quarto Report** rendert eine `.qmd`-Datei über `quarto::quarto_render()` in ein oder mehrere Ausgabeformate. Voreingestellt ist das zuletzt erzeugte Paket; über die Schaltfläche **Browse** können Sie jedoch eine beliebige `.qmd`-Datei auswählen. Drei Formate stehen zur Verfügung:

- **HTML** — in sich geschlossenes HTML-Dokument. Funktioniert auf allen Plattformen ohne zusätzliche Software. Empfohlene Vorgabe.
- **Word (.docx)** — gerendert mit der mitgelieferten Formatvorlage `reference.docx`. Geeignet zur Bearbeitung und Weitergabe.
- **PDF** — **erfordert eine LaTeX-Installation** (siehe Systemanforderungen weiter unten). Wird kein LaTeX erkannt, ist die PDF-Option in den Formatkästchen automatisch ausgeblendet.

Die einstufigen Renderer früherer Versionen, `render_gam_report()` und `export_gam_docx()`, stehen für die programmgesteuerte Nutzung weiterhin zur Verfügung.

### 13.3 Inhalt des Berichts

Der erzeugte Bericht enthält:

- **Datensatz & Modell:** Datendatei, Stichprobenumfang, Familie, Verfahren und Angaben zur Anpassung
- **Ergebniszusammenfassung:**  $R^2$ ,  $CV\text{-}R^2$ , erklärte Devianz, AIC und verwandte Gütemaße
- **Modellgleichung:** Modellgleichung mit  $f_i$ -Notation für die Glättungsterme und expliziten Koeffizienten für die linearen Terme, dazu die Definitionen der glatten Funktionen. Lange Gleichungen werden im PDF über mehrere Zeilen umbrochen.
- **Zusammenfassungstabellen:** Glättungsterme (EDF, F, p-Wert) und parametrische Terme (Schätzwert, Standardfehler, t, p-Wert)
- **Aktivierte Interaktionen:** eine Matrix der im Modell enthaltenen Tensor- und Faktor-Glättungsterm-Interaktionen (wird nur angezeigt, wenn Interaktionen vorhanden sind)
- **Variable Importance:** Balkendiagramm, geordnet nach F-Statistik (Glättungsterme) oder  $|t|$  (parametrisch)
- **Glatte Effekte** (alle Termtypen):
  - Diagramme univariater Glättungsterme (statische ggplot2-Fassungen der interaktiven Diagramme)

- Mehrlinien-Diagramme der Faktor-Glättungsterme
- Heatmaps der Interaktionen (*ti*- und *te*-Terme)
- Diagramme parametrischer Terme (Streudiagramm für numerische, Balkendiagramm für faktorielle Terme)
- Heatmap der **Prädiktorkorrelationen**
- **Diagnostik:** Residualdiagnostik und Streudiagramm beobachtet gegen prognostiziert
- **mgcv-Zusammenfassung:** die vollständige Textausgabe des angepassten Modells
- **Anhang zur Notation:** ein unnummerierter Schlussabschnitt, der die im Bericht verwendete Notation der Wertbeitragsfunktion erläutert

Nach erfolgreicher Erzeugung erscheint ein weißes Häkchen auf der Schaltfläche Generate Quarto Report.

## 14 Export von Modellfunktionen

mgcvUI kann die glatten Funktionen des angepassten GAM als eigenständigen Code in mehreren Sprachen exportieren. Jeder Glättungsterm wird auf einem Gitter mit 200 Punkten ausgewertet und als Nachschlagetabelle mit linearer Interpolation exportiert. So lässt sich das Modell außerhalb von R nutzen, ohne von `mgcv` abzuhängen.

Der Funktionsexport erfolgt in der aktuellen Version programmgesteuert. Die exportierenden Hilfsfunktionen nehmen ein angepasstes Ergebnisobjekt entgegen (erzeugt von `fit_gam()`) und schreiben die unten beschriebenen Dateien:

- `generate_r_code()`, `generate_python_code()` und `generate_cpp_code()` — Quelldateien je Sprache
- `export_functions_sqlite()` — SQLite-Datenbank mit Nachschlagetabellen
- `export_functions_zip()` — bündelt eine beliebige Kombination der vier Formate in einem ZIP-Archiv

### 14.1 R-Funktionen

Datei: `gam_functions.R`

- Kopfkommentare mit Zielvariable,  $R^2$ ,  $CV\text{-}R^2$  und der Formel.
- Eine Konstante `INTERCEPT`.
- Je Glättungsvariable eine `approxfun()` mit vorab berechneten  $x/y$ -Vektoren.
- Koeffizienten der linearen Terme als benannte Konstanten.
- Verwendung: Achsenabschnitt, jedes `g_var(value)` und jedes `coef_var * value` aufsummieren.

### 14.2 Python-Funktionen

Datei: `gam_functions.py`

- `numpy`-Import.
- Konstante `INTERCEPT`.
- Je Variable: Arrays `_x_var` und `_fx_var` sowie eine Funktion `g_var(x)` auf Basis von `np.interp()`.
- Koeffizienten der linearen Terme als Konstanten `COEF_*`.

### 14.3 C++-Header

Datei: `gam_functions.hpp`

- Reine Header-Datei mit `#pragma once`.
- Namensraum `gam_functions`.
- Hilfsfunktion `interp()` auf Basis von `std::lower_bound`.
- Je Variable: Inline-Funktion mit statischen `const`-Arrays.
- Koeffizienten der linearen Terme als `constexpr double`.

### 14.4 SQLite-Datenbank

Datei: `gam_functions.sqlite`

- Tabelle `smooth_grids`: Spalten (`variable`, `x`, `fx`) — die Nachschlagetabellen mit 200 Punkten.
- Tabelle `model_info`: `response`, `r_squared`, `cv_rsqr`, `dev_explained`, `aic`, `n_obs`, `family`, `method`, `formula`.

- Tabelle `linear_terms`: variable, coefficient.
- Tabelle `intercept`: Wert des Achsenabschnitts.

## 15 Vergleich mit earthUI und glmnetUI

Die Anwendung earth und ihre Post-Processing-Module bilden ein zusammengehöriges Werkzeugpaket für die Regressionsmodellierung. Sie teilen dasselbe Datenformat, dieselben Special-Spaltentypen, denselben RCA-Ablauf und dieselben Demodatensätze, verwenden jedoch unterschiedliche Modellierungs-Engines.

### 15.1 Wesentliche Unterschiede

| Merkmal            | mgcvUI                                    | earthUI                   | glmnetUI               |
|--------------------|-------------------------------------------|---------------------------|------------------------|
| Verfahren          | GAM (mgcv)                                | MARS/Earth (earth)        | Elastic Net (glmnet)   |
| Zusammenhänge      | Glatt nichtlinear                         | Stückweise linear         | Linear                 |
| Variablenselektion | Schrumpfung über <code>select=TRUE</code> | Schrittweise Beschneidung | Lasso-Strafterm        |
| Interaktionen      | Tensorprodukte (te/ti)                    | Hinge-Produkte (Grad 2–3) | Paarweise Produkte     |
| Partielle Effekte  | Glatte Kurven mit Konfidenzintervall      | g-Funktionen (stückweise) | Lineare Steigungen     |
| Earth-Pipeline     | Importiert Knoten aus earth               | Quelle der Knoten         | Importiert Hinge-Basis |
| Code-Export        | R, Python, C++, SQLite                    | R, Python, C++, SQLite    | —                      |
| Berichtsformate    | Word, PDF, HTML                           | Word, PDF, HTML           | Word, PDF, HTML        |

Ein Alleinstellungsmerkmal von earthUI verdient Erwähnung: earthUI enthält eine experimentelle Prolog/DCG-Funktion, die strukturierte Objektmerkmale aus MLS-Freitextbemerkungen extrahiert.

### 15.2 Gemeinsame Funktionen

Alle drei Werkzeuge bieten:

- einen gemeinsamen `regProj`-Projektbaum (Vorgabe `~/regProj`) — Projekte, Eingabedateien und verfahrensspezifische Ausgaben werden von den Schwesternanwendungen gemeinsam genutzt
- CSV- und Excel-Import mit Umwandlung der Spaltennamen in `snake_case`
- Locale-Unterstützung für über 30 Länder mit CSV-Trennzeichen, Dezimaltrennzeichen und Papierformat
- besondere Spaltenkennzeichnungen (`contract_date`, `living_area`, `area`, `latitude`, `longitude` usw.)
- automatische Berechnung von `sale_age` aus dem Stichtag
- Umschalter für Dark/Light Mode (Nord-Theme)
- dauerhafte Speicherung der Einstellungen (SQLite und/oder `localStorage`)
- Korrelations-Heatmap, Variable Importance, Diagnostik
- CQA-Bewertung und RCA-Anpassungsablauf
- Vergleichswertraster mit Auswahl der Vergleichsobjekte und formatierter Excel-Ausgabe
- Kompatibilität mit denselben Demodatensätzen (`Appraisal_1.csv`)

### 15.3 Empfohlener Arbeitsablauf

- **Beginnen Sie mit earthUI**, um die wichtigen Variablen, nichtlinearen Zusammenhänge und Interaktionen aufzuspüren. Der automatische Modellaufbau von earth liefert den

besten Ausgangspunkt.

- **Verfeinern Sie mit mgcvUI**, wenn Sie glattere partielle Effekte wünschen. Importieren Sie die Knoten von earthUI für einen nahtlosen Übergang.
- **Verwenden Sie glmnetUI**, wenn Sie ein rein lineares Modell mit konsequenter Variablenselektion oder mit Vorzeichenbeschränkungen für die Koeffizienten benötigen.
- **Vergleichen Sie alle drei** an denselben Daten, um zu beurteilen, ob die zusätzliche Komplexität nichtlinearer Modelle gerechtfertigt ist.

## 16 Demodatensatz: Appraisal\_1.csv

### 16.1 Beschreibung

Der Demodatensatz wird gemeinsam mit earthUI und glmnetUI genutzt. Er liegt im Ordner demo\_mls/ des mgcvUI-Quelltextes oder, falls earthUI installiert ist:

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

Die Datei enthält rund 1.500 Verkaufsdatensätze von Wohnimmobilien aus einem simulierten MLS-Export, mit dem Bewertungsobjekt in Zeile 1 und den Vergleichsverkäufen in den übrigen Zeilen. Die Daten bilden Verkäufe von Einfamilienhäusern in einem Markt mit mehreren Gebieten ab, mit einer Bandbreite an Objektgrößen, Baualtern und Lagen.

Es handelt sich nicht um echte Daten, sie beruhen jedoch auf einer realistischen Nachbarschaft in Nordkalifornien. Sämtliche identifizierenden Angaben wurden verändert oder entfernt.

### 16.2 Spalten

| Spalte              | Typ       | Special Type  | Beschreibung                                                                          |
|---------------------|-----------|---------------|---------------------------------------------------------------------------------------|
| sale_type           | character | sale_type     | Datensatztyp-Code, der das Bewertungsobjekt von den Vergleichsverkäufen unterscheidet |
| weight              | numeric   | weight        | Beobachtungsgewicht (0 = von der Anpassung ausschließen)                              |
| id                  | numeric   | display_only  | Interne Datensatzkennung                                                              |
| property_id         | numeric   | display_only  | MLS-Objektkennung                                                                     |
| listing_id          | character | display_only  | MLS-Angebotsnummer                                                                    |
| parcel_number       | character | display_only  | Flurstücksnummer der Bewertungsbehörde (APN)                                          |
| street_address      | character | display_only  | Objektadresse                                                                         |
| city_name           | character | display_only  | Stadt                                                                                 |
| postal_code         | character | display_only  | Postleitzahl                                                                          |
| county_name         | character | display_only  | Landkreis                                                                             |
| contract_date       | Date      | contract_date | Kaufvertragsdatum (berechnet sale_age)                                                |
| sale_age            | numeric   | —             | Tage vom Vertragsdatum bis zum Stichtag (vorab berechnet)                             |
| coe_date            | Date      | display_only  | Datum des Abschlusses der Treuhandabwicklung                                          |
| listing_status      | character | display_only  | Angebotsstatus (z. B. „Sold“)                                                         |
| sale_price          | numeric   | (Zielgröße)   | Verkaufspreis — Zielvariable                                                          |
| rent                | numeric   | —             | Monatsmiete (für Modelle mit mehreren Zielgrößen)                                     |
| list_price          | numeric   | display_only  | Angebotspreis                                                                         |
| original_list_price | numeric   | display_only  | Ursprünglicher Angebotspreis                                                          |
| living_sqft         | numeric   | living_area   | Bruttowohnfläche in Quadratfuß                                                        |
| ppsf                | numeric   | display_only  | Preis je Quadratfuß                                                                   |
| beds_total          | integer   | —             | Anzahl der Schlafzimmer                                                               |
| baths_total         | numeric   | —             | Gesamtzahl der Bäder (z. B. 2.5 = 2 volle + 1 halbes)                                 |
| lot_size            | numeric   | lot_size      | Grundstücksgröße in Quadratfuß                                                        |
| area_id             | integer   | area          | MLS-Gebietskennung                                                                    |
| area_text           | character | display_only  | Gebietsname                                                                           |
| age                 | numeric   | actual_age    | Objektalter in Jahren                                                                 |
| year_built          | integer   | display_only  | Baujahr                                                                               |
| latitude            | numeric   | latitude      | Breitengrad (für das Modell auf 3 Nachkommastellen gerundet)                          |
| longitude           | numeric   | longitude     | Längengrad (für das Modell auf 3 Nachkommastellen gerundet)                           |
| latitude6           | numeric   | display_only  | Breitengrad in voller Genauigkeit                                                     |
| longitude6          | numeric   | display_only  | Längengrad in voller Genauigkeit                                                      |
| garage_spaces       | integer   | —             | Anzahl der Garagenstellplätze                                                         |

| Spalte                        | Typ       | Special Type              | Beschreibung                                     |
|-------------------------------|-----------|---------------------------|--------------------------------------------------|
| <code>fp_count</code>         | integer   | —                         | Anzahl der Kamine                                |
| <code>no_of_stories</code>    | numeric   | —                         | Anzahl der Geschosse                             |
| <code>style</code>            | character | —                         | Architekturstil                                  |
| <code>view</code>             | character | —                         | Art der Aussicht (z. B. „Neighborhood“, „Hills“) |
| <code>days_on_market</code>   | integer   | <code>dom</code>          | Vermarktungsdauer in Tagen                       |
| <code>listing_date</code>     | Date      | <code>listing_date</code> | Angebotsdatum                                    |
| <code>sale_concessions</code> | numeric   | <code>concessions</code>  | Verkäuferzugeständnisse                          |

### 16.3 Vorgeschlagener Schnelleinstieg

1. Starten Sie die mgcv-Routine: `earthUI::launch_mgcv()`
2. Legen Sie ein neues Projekt (Abschnitt 1) mit dem Purpose **For Appraisal** an
3. Legen Sie `Appraisal_1.csv` in den Eingabeordner des Projekts und wählen Sie die Datei in Abschnitt 2 aus (klicken Sie bei Bedarf auf **Refresh**)
4. Wählen Sie `sale_price` als Zielgröße
5. Setzen Sie Response Transform optional auf **Log (natural)**, um proportionale Anpassungen zu erhalten
6. Weisen Sie die Special Types wie in der Tabelle oben zu
7. Beziehen Sie folgende Prädiktoren ein: `sale_age`, `living_sqft`, `baths_total`, `lot_size`, `area_id` (als Faktor), `age`, `latitude`, `longitude`, `garage_spaces`
8. Behalten Sie die Vorgaben bei: REML,  $\gamma = 1.2$ , tp-Basis,  $k = \text{auto}$ , `select = TRUE`
9. Klicken Sie auf **Fit Mgcv GAM Model**
10. Prüfen Sie die Registerkarten Summary, Contribution (Diagramme der Glättungsterme) und Diagnostics
11. Laden Sie die Ausgabe herunter (Schritt 7) und prüfen Sie die CQA-Rangfolge
12. Berechnen Sie die RCA-Anpassungen (Schritt 8) mit einem CQA-Wert von etwa 5.00
13. Erzeugen Sie ein Vergleichswertraster (Schritt 9) mit den empfohlenen Vergleichsobjekten

## 17 Systemanforderungen & Fehlerbehebung

### 17.1 Unterstützte Plattformen

mgcvUI läuft unter **macOS**, **Windows** und **Linux**. Die Anwendung wird unter macOS mit RStudio entwickelt und vorrangig getestet. Plattformspezifische Hinweise finden Sie weiter unten.

### 17.2 R-Version

Erforderlich ist  $R \geq 4.1.0$ . RStudio Desktop (2023.06 oder neuer) wird dringend empfohlen — es bringt **Quarto** und **pandoc** mit (für das Rendern der Berichte erforderlich) und bietet eine bequeme Umgebung zum Starten der Anwendung.

### 17.3 Erforderliche R-Pakete

Die folgenden Pakete werden automatisch als Abhängigkeiten installiert:

| Paket             | Zweck                                                                   |
|-------------------|-------------------------------------------------------------------------|
| shiny             | Framework für Webanwendungen                                            |
| mgcv              | Anpassung der GAM-Modelle                                               |
| valengrCore       | Gemeinsamer Wertermittlungskern (Special-Rollen, Übernahme aus earthUI) |
| ggplot2, plotly   | Statische und interaktive Beitragsdiagramme                             |
| gratia            | Extraktion der Glättungsterme und Diagnostik                            |
| DT                | Interaktive Datentabellen                                               |
| quarto, rmarkdown | Erzeugung und Rendern der Quarto-Berichte                               |
| readxl            | Import von Excel-Dateien                                                |
| shinyFiles        | Datei- und Ordnerauswahl                                                |
| RSQLite, DBI      | Einstellungs- und Projektdatenbanken (SQLite)                           |
| jsonlite          | Serialisierung des Variablenzustands                                    |

## 17.4 Optionale Abhängigkeiten

| Komponente                                 | Wann erforderlich                                                                                                                                                                                           |
|--------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>LaTeX</b> (TinyTeX, MiKTeX oder MacTeX) | Nur für die Erzeugung von PDF-Berichten. Wird keine LaTeX-Installation erkannt, ist die PDF-Option im Formatauswahlfeld automatisch ausgeblendet. Installation mit: <code>tinytex::install_tinytex()</code> |
| <b>sysfonts + showtext</b>                 | Schriftart Roboto Condensed für ggplot2-Diagramme. Ist sie nicht verfügbar oder besteht keine Internetverbindung, weicht die Anwendung automatisch auf die serifenlose System-schrift aus.                  |
| <b>earth</b> ( $\geq 5.3.0$ )              | Nur für die Importpipeline aus earthUI erforderlich. Für den eigenständigen Betrieb nicht nötig.                                                                                                            |
| <b>callr</b>                               | Modellanpassung im Hintergrund (hält die Anwendung bedienbar). Ohne dieses Paket läuft die Anpassung synchron.                                                                                              |
| <b>writexl, openxlsx</b>                   | Excel-Downloads (Modellausgabe, RCA, Vergleichswertraster). Die Anwendung weist auf ein Fehlen hin.                                                                                                         |
| <b>bslib</b>                               | Bootstrap-5-Theming (Nord Light/Dark).                                                                                                                                                                      |
| <b>officer</b>                             | Programmgesteuerter Word-Export früherer Versionen ( <code>export_gam_docx()</code> ).                                                                                                                      |

## 17.5 Plattformspezifische Hinweise

### 17.5.1 macOS

Die wenigsten Probleme. Für Systembibliotheken wird Homebrew empfohlen. Werden PDF-Berichte benötigt:

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

### 17.5.2 Windows

Funktioniert gut mit RStudio Desktop. Wichtige Hinweise:

- **Firmen- oder gesperrte Rechner:** Ist das Temp-Verzeichnis oder das Benutzerdatenverzeichnis eingeschränkt, können die dauerhafte Speicherung der Einstellungen und die Berichtserzeugung fehlschlagen. Die Anwendung zeigt dann eine Warnung in der R-Konsole an, arbeitet aber ohne Speicherung weiter.
- **LaTeX:** Installieren Sie TinyTeX aus R heraus (gleicher Befehl wie oben für macOS) oder MiKTeX von [miktex.org](http://miktex.org).
- **Lange Dateipfade:** Windows begrenzt Pfade auf 260 Zeichen. Halten Sie den Pfad Ihres Arbeitsverzeichnisses kurz.

### 17.5.3 Linux

Am uneinheitlichsten zwischen den Distributionen. Unter Ubuntu/Debian sind unter Umständen Systembibliotheken erforderlich, bevor sich die R-Pakete kompilieren lassen:

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

Für PDF-Berichte: `tinytex::install_tinytex()` oder `texlive-full` über die Paketverwaltung des Systems installieren.

Auf Servern ohne Anzeige (headless) werden interaktive Plotly-Diagramme in Berichten als statische Alternativen gerendert. Die Shiny-Anwendung selbst benötigt eine Verbindung über einen Webbrowser.

## 17.6 Verhalten bei fehlenden Komponenten

mgcvUI ist so ausgelegt, dass es auch bei fehlenden optionalen Komponenten arbeitet:

| Fehlende Komponente                           | Verhalten                                                                                                                                                      |
|-----------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Kein LaTeX                                    | Die PDF-Option ist im Auswahlfeld der Berichtsformate ausgeblendet. HTML und Word bleiben verfügbar.                                                           |
| Kein Internet / Schriftarten fehlen           | Roboto Condensed wird durch die serifenlose Systemschrift ersetzt. Eine Konsolenmeldung wird protokolliert.                                                    |
| Kein RSQlite / schreibgeschütztes Dateisystem | Einstellungen bleiben zwischen Sitzungen nicht erhalten. Eine Warnung erscheint in der Konsole. Ansonsten läuft die Anwendung normal.                          |
| gratia::appraise schlägt fehl                 | Die Registerkarte Diagnostics und die Berichte zeigen eine Platzhaltermeldung. Verwenden Sie ersatzweise <code>mgcv::gam.check()</code> in der Konsole.        |
| Temp-Verzeichnis nicht beschreibbar           | Die Berichtserzeugung schlägt mit einer klaren Fehlermeldung fehl. Abhilfe: Setzen Sie die Umgebungsvariable <code>TMPDIR</code> auf einen beschreibbaren Ort. |

## 17.7 Fehlerbehebung

### 17.7.1 „PDF option is not available“

Es wurde keine LaTeX-Installation erkannt. Führen Sie in R aus:

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

Starten Sie danach die Anwendung neu. Die PDF-Option erscheint dann.

### 17.7.2 „Settings will not persist“

Das Verzeichnis der SQLite-Datenbank kann nicht angelegt oder beschrieben werden. Prüfen Sie die Berechtigungen für:

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

### 17.7.3 „Diagnostics plot unavailable“

Die Funktion `gratia::appraise()` schlägt bei komplexen Modellen (viele Tensor-Interaktionen, nicht standardmäßige Familien) gelegentlich fehl. Dies ist ein bekanntes Problem der Ursprungsbibliothek. Abhilfe: Führen Sie `mgcv::gam.check(model)` in der R-Konsole aus, um die Diagnostikdiagramme unmittelbar anzuzeigen.

### 17.7.4 „Report error: number of items to replace...“

Dies war ein bekanntes Problem früherer Versionen im Zusammenhang mit dem Paket `gratia`. Stellen Sie sicher, dass Sie die neueste Fassung von `mgcvUI` installiert haben. Die aktuelle Fassung fängt diesen Fehler ab und erzeugt den Bericht mit einem Platzhalter für den Diagnostikabschnitt.

### 17.7.5 Port 7880 bereits belegt

Startet die Anwendung nicht, weil Port 7880 (aus einer früheren Sitzung) belegt ist, führen Sie aus:

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

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

### 17.7.6 Paketinstallation schlägt unter Linux fehl

Es fehlen Systembibliotheken. Installieren Sie die im Linux-Abschnitt oben aufgeführten Entwicklungsheder und wiederholen Sie `install.packages()`.

## Literatur

- Wood, S. N. (2003). „Thin-plate regression splines“. In: *Journal of the Royal Statistical Society (B)* 65.1, S. 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“. In: *Journal of the American Statistical Association* 99.467, S. 673–686. DOI: [10.1198/016214504000000980](https://doi.org/10.1198/016214504000000980).
- (2011). „Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models“. In: *Journal of the Royal Statistical Society (B)* 73.1, S. 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. Aufl. Chapman und Hall/CRC.
- Wood, S. N., N. Pya und B. Säfken (2016). „Smoothing parameter and model selection for general smooth models (with discussion)“. In: *Journal of the American Statistical Association* 111, S. 1548–1575. DOI: [10.1080/01621459.2016.1180986](https://doi.org/10.1080/01621459.2016.1180986).